

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/00	A2	(11) International Publication Number: WO 99/61599 (43) International Publication Date: 2 December 1999 (02.12.99)
(21) International Application Number: PCT/US99/11814 (22) International Filing Date: 27 May 1999 (27.05.99) (30) Priority Data: 60/087,080 28 May 1998 (28.05.98) US 09/141,908 28 August 1998 (28.08.98) US 60/100,880 22 September 1998 (22.09.98) US 60/119,139 8 February 1999 (08.02.99) US (71) Applicant: KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Place, Hayward, CA 94545 (US). (72) Inventors: ASHLEY, Gary; 1102 Verdemar Drive, Alameda, CA 94502 (US). BETLACH, Melanie, C.; Apartment #9, 11211 Bellevue Drive, Burlingame, CA 94010 (US). BETLACH, Mary; 2530 Diamond Street, San Francisco, CA 94131 (US). MCDANIEL, Robert; 698 Matadero Avenue, Palo Alto, CA 94306 (US). TANG, Li; 574 Cutwater Lane, Foster City, CA 94404 (US). (74) Agents: MURASHIGE, Kate et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).		(81) Designated States: AU, CA, IL, JP, MX, NZ, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE (57) Abstract Recombinant DNA compounds that encode all or a portion of the narbonolide polyketide synthase are used to express recombinant polyketide synthase genes in host cells for the production of narbonolide, narbonolide derivatives, and polyketides that are useful as antibiotics and as intermediates in the synthesis of compounds with pharmaceutical value.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 15/52, C12N 1/21, C12N 9/00, C12N 15/76	A3	(11) International Publication Number: WO 99/61599 (43) International Publication Date: 02 December 1999 (02.12.1999)					
<table border="0"><tr><td data-bbox="207 456 1081 499">(21) International Application Number: PCT/US99/11814</td><td data-bbox="1102 499 1921 556" rowspan="4">Published</td></tr><tr><td data-bbox="207 528 1081 585">(22) International Filing Date: 27 May 1999 (27.05.1999)</td></tr><tr><td data-bbox="207 614 1081 813">(30) Priority Data: 09/141,908 28 August 1998 (28.08.1998) US 60/087,080 28 May 1998 (28.05.1998) US 60/100,880 22 September 1998 (22.09.1998) US 60/119,139 08 February 1999 (08.02.1999) US</td></tr><tr><td data-bbox="207 842 1081 1071">(60) Parent Application or Grant KOSAN BIOSCIENCES, INC. [/]; (). ASHLEY, Gary [/]; (). BETLACH, Melanie, C. [/]; (). BETLACH, Mary [/]; (). MCDANIEL, Robert [/]; (). TANG, Li [/]; (). MURASHIGE, Kate; ().</td></tr></table>			(21) International Application Number: PCT/US99/11814	Published	(22) International Filing Date: 27 May 1999 (27.05.1999)	(30) Priority Data: 09/141,908 28 August 1998 (28.08.1998) US 60/087,080 28 May 1998 (28.05.1998) US 60/100,880 22 September 1998 (22.09.1998) US 60/119,139 08 February 1999 (08.02.1999) US	(60) Parent Application or Grant KOSAN BIOSCIENCES, INC. [/]; (). ASHLEY, Gary [/]; (). BETLACH, Melanie, C. [/]; (). BETLACH, Mary [/]; (). MCDANIEL, Robert [/]; (). TANG, Li [/]; (). MURASHIGE, Kate; ().
(21) International Application Number: PCT/US99/11814	Published						
(22) International Filing Date: 27 May 1999 (27.05.1999)							
(30) Priority Data: 09/141,908 28 August 1998 (28.08.1998) US 60/087,080 28 May 1998 (28.05.1998) US 60/100,880 22 September 1998 (22.09.1998) US 60/119,139 08 February 1999 (08.02.1999) US							
(60) Parent Application or Grant KOSAN BIOSCIENCES, INC. [/]; (). ASHLEY, Gary [/]; (). BETLACH, Melanie, C. [/]; (). BETLACH, Mary [/]; (). MCDANIEL, Robert [/]; (). TANG, Li [/]; (). MURASHIGE, Kate; ().							
(54) Title: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE (54) Titre: SYNTHASE DE POLYKETIDE NARBONOLIDE DE RECOMBINAISON (57) Abstract <p>Recombinant DNA compounds that encode all or a portion of the narbonolide polyketide synthase are used to express recombinant polyketide synthase genes in host cells for the production of narbonolide, narbonolide derivatives, and polyketides that are useful as antibiotics and as intermediates in the synthesis of compounds with pharmaceutical value.</p> (57) Abrégé <p>L'invention porte sur des composés d'ADN de recombinaison qui codent tout ou partie de la synthase de polyketide narbonolide et qui sont utilisés pour exprimer des gènes de la synthase de polyketides de recombinaison dans des cellules hôtes pour la production de narbonolide et de dérivés de narbonolide, et de polyketides qui sont utiles comme antibiotiques et comme intermédiaires dans la synthèse de composés ayant une valeur pharmaceutique.</p>							

RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

Reference to Government Funding

5 This invention was supported in part by SBIR grant 1R43-CA75792-01. The U.S. government has certain rights in this invention.

Field of the Invention

10 The present invention provides recombinant methods and materials for producing polyketides by recombinant DNA technology. More specifically, it relates to narbonolides and derivatives thereof. The invention relates to the fields of agriculture, animal husbandry, chemistry, medicinal chemistry, medicine, molecular biology, pharmacology, and veterinary technology.

Background of the Invention

15 Polyketides represent a large family of diverse compounds synthesized from 2-carbon units through a series of condensations and subsequent modifications. Polyketides occur in many types of organisms, including fungi and mycelial bacteria, in particular, the actinomycetes. There is a wide variety of polyketide structures, and the class of polyketides
20 encompasses numerous compounds with diverse activities. Tetracycline, erythromycin, FK506, FK520, narbomycin, picromycin, rapamycin, spinocyn, and tylosin, are examples of such compounds. Given the difficulty in producing polyketide compounds by traditional chemical methodology, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce
25 polyketide compounds. See PCT publication Nos. WO 93/13663; WO 95/08548; WO 96/40968; WO 97/02358; and WO 98/27203; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; and 5,712,146; Fu *et al.*, 1994, *Biochemistry* 33: 9321-9326; McDaniel *et al.*, 1993, *Science* 262: 1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

30 Polyketides are synthesized in nature by polyketide synthase (PKS) enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. PKS enzymes are encoded by PKS genes that usually consist of three or more open reading frames (ORFs).

Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

Modular PKSs are responsible for producing a large number of 12, 14, and 16-membered macrolide antibiotics including methymycin, erythromycin, narbomycin, picromycin, and tylosin. These large multifunctional enzymes (>300,000 kDa) catalyze the biosynthesis of polyketide macrolactones through multistep pathways involving decarboxylative condensations between acyl thioesters followed by cycles of varying β -carbon processing activities (see O'Hagan, D. *The polyketide metabolites*; E. Horwood: New York, 1991, incorporated herein by reference). The modular PKS are generally encoded in multiple ORFs. Each ORF typically comprises two or more "modules" of ketosynthase activity, each module of which consists of at least two (if a loading module) and more typically three or more enzymatic activities or "domains."

During the past half decade, the study of modular PKS function and specificity has been greatly facilitated by the plasmid-based *Streptomyces coelicolor* expression system developed with the 6-deoxyerythronolide B (6-dEB) synthase (DEBS) genes (see Kao *et al.*, 1994, *Science*, 265: 509-512, McDaniel *et al.*, 1993, *Science* 262: 1546-1557, and U.S. Patent Nos. 5,672,491 and 5,712,146, each of which is incorporated herein by reference). The advantages to this plasmid-based genetic system for DEBS were that it overcame the tedious and limited techniques for manipulating the natural DEBS host organism, *Saccharopolyspora erythraea*, allowed more facile construction of recombinant PKSs, and reduced the complexity of PKS analysis by providing a "clean" host background. This system also expedited construction of the first combinatorial modular polyketide library in *Streptomyces* (see PCT publication No. WO 98/49315, incorporated herein by reference).

The ability to control aspects of polyketide biosynthesis, such as monomer selection and degree of β -carbon processing, by genetic manipulation of PKSs has stimulated great interest in the combinatorial engineering of novel antibiotics (see Hutchinson, 1998, *Curr. Opin. Microbiol.* 1: 319-329; Carreras and Santi, 1998, *Curr. Opin. Biotech.* 9: 403-411; and U.S. Patent Nos. 5,712,146 and 5,672,491, each of which is incorporated herein by reference). This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not

produce the polyketide. The technology also allows one to produce molecules that are structurally related to, but distinct from, the polyketides produced from known PKS gene clusters. It has been possible to manipulate modular PKS genes other than the narbonolide PKS using generally known recombinant techniques to obtain altered and hybrid forms. See, 5 e.g., U.S. Patent Nos. 5,672,491 and 5,712,146 and PCT publication No. WO 98/49315. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units" *Biochemistry* 38(5):1643-1651, and Gokhale *et al.*, 16 Apr. 1999, Dissecting and Exploiting Intermodular Communication in Polyketide Synthases", *Science* 284: 482-485.

10 The present invention provides methods and reagents relating to the modular PKS gene cluster for the polyketide antibiotics known as narbomycin and picromycin. Narbomycin is produced in *Streptomyces narbonensis*, and both narbomycin and picromycin are produced in *S. venezuelae*. These species are unique among macrolide producing organisms in that they produce, in addition to the 14-membered macrolides narbomycin and 15 picromycin (picromycin is shown in Figure 1, compound 1), the 12-membered macrolides neomethymycin and methymycin (methymycin is shown in Figure 1, compound 2). Narbomycin differs from picromycin only by lacking the hydroxyl at position 12. Based on the structural similarities between picromycin and methymycin, it was speculated that methymycin would result from premature cyclization of a hexaketide intermediate in the 20 picromycin pathway.

Glycosylation of the C5 hydroxyl group of the polyketide precursor, narbonolide, is achieved through an endogenous desosaminyl transferase to produce narbomycin. In *Streptomyces venezuelae*, narbomycin is then converted to picromycin by the endogenously produced narbomycin hydroxylase. (See Figure 1) Thus, as in the case of other macrolide 25 antibiotics, the macrolide product of the narbonolide PKS is further modified by hydroxylation and glycosylation. Figure 1 also shows the metabolic relationships of the compounds discussed above.

Picromycin (Figure 1, compound 1) is of particular interest because of its close structural relationship to ketolide compounds (e.g. HMR 3004, Figure 1, compound 3). The 30 ketolides are a new class of semi-synthetic macrolides with activity against pathogens resistant to erythromycin (see Agouridas *et al.*, 1998, *J. Med. Chem.* 41: 4080-4100, incorporated herein by reference). Thus, genetic systems that allow rapid engineering of the narbonolide PKS would be valuable for creating novel ketolide analogs for pharmaceutical

applications. Furthermore, the production of picromycin as well as novel compounds with useful activity could be accomplished if the heterologous expression of the narbonolide PKS in *Streptomyces lividans* and other host cells were possible. The present invention meets these and other needs.

5

Disclosure of the Invention

The present invention provides recombinant methods and materials for expressing PKSs derived in whole and in part from the narbonolide PKS and other genes involved in narbomycin and picromycin biosynthesis in recombinant host cells. The invention also
10 provides the polyketides derived from the narbonolide PKS. The invention provides the complete PKS gene cluster that ultimately results, in *Streptomyces venezuelae*, in the production of picromycin. The ketolide product of this PKS is narbonolide. Narbonolide is glycosylated to obtain narbomycin and then hydroxylated at C12 to obtain picromycin. The enzymes responsible for the glycosylation and hydroxylation are also provided in
15 recombinant form by the invention.

Thus, in one embodiment, the invention is directed to recombinant materials that contain nucleotide sequences encoding at least one domain, module, or protein encoded by a narbonolide PKS gene. The recombinant materials may be "isolated." The invention also provides recombinant materials useful for conversion of ketolides to antibiotics. These
20 materials include recombinant DNA compounds that encode the C12hydroxylase (the *picK* gene), the desosamine biosynthesis and desosaminyl transferase enzymes, and the beta-glucosidase enzyme involved in picromycin biosynthesis in *S. venezuelae* and the recombinant proteins that can be produced from these nucleic acids in the recombinant host cells of the invention.

25 In one embodiment, the invention provides a recombinant expression system that comprises a heterologous promoter positioned to drive expression of the narbonolide PKS, including a "hybrid" narbonolide PKS.. In a preferred embodiment, the promoter is derived from a PKS gene. In a related embodiment, the invention provides recombinant host cells comprising the vector that produces narbonolide. In a preferred embodiment, the host cell is
30 *Streptomyces lividans* or *S. coelicolor*.

In another embodiment, the invention provides a recombinant expression system that comprises the desosamine biosynthetic genes as well as the desosaminyl transferase gene. In a related embodiment, the invention provides recombinant host cells comprising a vector that

produces the desosamine biosynthetic gene products and desosaminyl transferase gene product. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In another embodiment, the invention provides a method for desosaminylating polyketide compounds in recombinant host cells, which method comprises expressing the
5 PKS for the polyketide and the desosaminyl transferase and desosamine biosynthetic genes in a host cell. In a preferred embodiment, the host cell expresses a beta-glucosidase gene as well. This preferred method is especially advantageous when producing desosaminylated polyketides in *Streptomyces* host cells, because such host cells typically glucosylate desosamine residues of polyketides, which can decrease desired activity, such as antibiotic
10 activity. By coexpression of beta-glucosidase, the glucose residue is removed from the polyketide.

In another embodiment, the invention provides the *picK* hydroxylase gene in recombinant form and methods for hydroxylating polyketides with the recombinant gene product. The invention also provides polyketides thus produced and the antibiotics or other
15 useful compounds derived therefrom.

In another embodiment, the invention provides a recombinant expression system that comprises a promoter positioned to drive expression of a "hybrid" PKS comprising all or part of the narbonolide PKS and at least a part of a second PKS, or comprising a narbonolide PKS modified by deletions, insertions and/or substitutions. In a related embodiment, the invention
20 provides recombinant host cells comprising the vector that produces the hybrid PKS and its corresponding polyketide. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In a related embodiment, the invention provides recombinant materials for the production of libraries of polyketides wherein the polyketide members of the library are
25 synthesized by hybrid PKS enzymes of the invention. The resulting polyketides can be further modified to convert them to other useful compounds, such as antibiotics, typically through hydroxylation and/or glycosylation. Modified macrolides provided by the invention that are useful intermediates in the preparation of antibiotics are of particular benefit.

In another related embodiment, the invention provides a method to prepare a nucleic
30 acid that encodes a modified PKS, which method comprises using the narbonolide PKS encoding sequence as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, insertion, or replacement. The thus modified narbonolide PKS encoding nucleotide sequence can then be expressed in a

suitable host cell and the cell employed to produce a polyketide different from that produced by the narbonolide PKS. In addition, portions of the narbonolide PKS coding sequence can be inserted into other PKS coding sequences to modify the products thereof. The narbonolide PKS can itself be manipulated, for example, by fusing two or more of its open reading
5 frames, particularly those for extender modules 5 and 6, to make more efficient the production of 14-membered as opposed to 12-membered macrolides.

In another related embodiment, the invention is directed to a multiplicity of cell colonies, constituting a library of colonies, wherein each colony of the library contains an expression vector for the production of a modular PKS derived in whole or in part from the
10 narbonolide PKS. Thus, at least a portion of the modular PKS is identical to that found in the PKS that produces narbonolide and is identifiable as such. The derived portion can be prepared synthetically or directly from DNA derived from organisms that produce narbonolide. In addition, the invention provides methods to screen the resulting polyketide and antibiotic libraries.

15 The invention also provides novel polyketides and antibiotics or other useful compounds derived therefrom. The compounds of the invention can be used in the manufacture of another compound. In a preferred embodiment, the antibiotic compounds of the invention are formulated in a mixture or solution for administration to an animal or human.

20 These and other embodiments of the invention are described in more detail in the following description, the examples, and claims set forth below.

Brief Description of the Figures

Figure 1 shows the structures of picromycin (compound 1), methymycin (compound
25 2), and the ketolide HMR 3004 (compound 3) and the relationship of several compounds related to picromycin.

Figure 2 shows a restriction site and function map of cosmid pKOS023-27.

Figure 3 shows a restriction site and function map of cosmid pKOS023-26.

Figure 4 has three parts. In Part A, the structures of picromycin (A(a)) and
30 methymycin (A(b)) are shown, as well as the related structures of narbomycin, narbonolide, and methynolide. In the structures, the bolded lines indicate the two or three carbon chains produced by each module (loading and extender) of the narbonolide PKS. Part B shows the organization of the narbonolide PKS genes on the chromosome of *Streptomyces venezuelae*,

including the location of the various module encoding sequences (the loading module domains are identified as sKS*, sAT, and sACP), as well as the *picB* thioesterase gene and two desosamine biosynthesis genes (*picCII* and *picCIII*). Part C shows the engineering of the *S. venezuelae* host of the invention in which the *picAI* gene has been deleted. In the Figure,
5 ACP is acyl carrier protein; AT is acyltransferase; DH is dehydratase; ER is enoylreductase; KR is ketoreductase; KS is ketosynthase; and TE is thioesterase.

Figure 5 shows the narbonolide PKS genes encoded by plasmid pKOS039-86, the compounds synthesized by each module of that PKS and the narbonolide (compound 4) and 10-deoxymethynolide (compound 5) products produced in heterologous host cells
10 transformed with the plasmid. The Figure also shows a hybrid PKS of the invention produced by plasmid pKOS038-18, which encodes a hybrid of DEBS and the narbonolide PKS. The Figure also shows the compound, 3,6-dideoxy-3-oxo-erythronolide B (compound 6), produced in heterologous host cells comprising the plasmid.

Figure 6 shows a restriction site and function map of plasmid pKOS039-104, which
15 contains the desosamine biosynthetic, beta-glucosidase, and desosaminyl transferase genes under transcriptional control of *actII-4*.

Modes of Carrying out the Invention

The present invention provides useful compounds and methods for producing
20 polyketides in recombinant host cells. As used herein, the term recombinant refers to a compound or composition produced by human intervention. The invention provides recombinant DNA compounds encoding all or a portion of the narbonolide PKS. The invention also provides recombinant DNA compounds encoding the enzymes that catalyze the further modification of the ketolides produced by the narbonolide PKS. The invention
25 provides recombinant expression vectors useful in producing the narbonolide PKS and hybrid PKSs composed of a portion of the narbonolide PKS in recombinant host cells. Thus, the invention also provides the narbonolide PKS, hybrid PKSs, and polyketide modification enzymes in recombinant form. The invention provides the polyketides produced by the recombinant PKS and polyketide modification enzymes. In particular, the invention provides
30 methods for producing the polyketides 10-deoxymethynolide, narbonolide, YC17, narbomycin, methymycin, neomethymycin, and picromycin in recombinant host cells.

To appreciate the many and diverse benefits and applications of the invention, the description of the invention below is organized as follows. First, a general description of

polyketide biosynthesis and an overview of the synthesis of narbonolide and compounds derived therefrom in *Streptomyces venezuelae* are provided. This general description and overview are followed by a detailed description of the invention in six sections. In Section I, the recombinant narbonolide PKS provided by the invention is described. In Section II, the recombinant desosamine biosynthesis genes, the desosaminyl transferase gene, and the beta-glucosidase gene provided by the invention are described. In Section III, the recombinant *picK* hydroxylase gene provided by the invention is described. In Section IV, methods for heterologous expression of the narbonolide PKS and narbonolide modification enzymes provided by the invention are described. In Section V, the hybrid PKS genes provided by the invention and the polyketides produced thereby are described. In Section VI, the polyketide compounds provided by the invention and pharmaceutical compositions of those compounds are described. The detailed description is followed by a variety of working examples illustrating the invention.

The narbonolide synthase gene, like other PKS genes, is composed of coding sequences organized in a loading module, a number of extender modules, and a thioesterase domain. As described more fully below, each of these domains and modules is a polypeptide with one or more specific functions. Generally, the loading module is responsible for binding the first building block used to synthesize the polyketide and transferring it to the first extender module. The building blocks used to form complex polyketides are typically acylthioesters, most commonly acetyl, propionyl, malonyl, methylmalonyl, and ethylmalonyl CoA. Other building blocks include amino acid like acylthioesters. PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between the acylthioester building blocks. Each module is responsible for binding a building block, performing one or more functions on that building block, and transferring the resulting compound to the next module. The next module, in turn, is responsible for attaching the next building block and transferring the growing compound to the next module until synthesis is complete. At that point, an enzymatic thioesterase activity cleaves the polyketide from the PKS. See, generally, Figure 5.

Such modular organization is characteristic of the modular class of PKS enzymes that synthesize complex polyketides and is well known in the art. The polyketide known as 6-deoxyerythronolide B is a classic example of this type of complex polyketide. The genes, known as *eryAI*, *eryAII*, and *eryAIII* (also referred to herein as the DEBS genes, for the proteins, known as DEBS1, DEBS2, and DEBS3, that comprise the 6-dEB synthase), that

code for the multi-subunit protein known as DEBS that synthesizes 6-dEB, the precursor polyketide to erythromycin, are described in U.S. Patent No. 5,824,513, incorporated herein by reference. Recombinant methods for manipulating modular PKS genes are described in U.S. Patent Nos. 5,672,491; 5,843,718; 5,830,750; and 5,712,146; and in PCT publication
5 Nos. WO 98/49315 and WO 97/02358, each of which is incorporated herein by reference.

The loading module of DEBS consists of two domains, an acyl-transferase (AT) domain and an acyl carrier protein (ACP) domain. Each extender module of DEBS, like those of other modular PKS enzymes, contains a ketosynthase (KS), AT, and ACP domains, and zero, one, two, or three domains for enzymatic activities that modify the beta-carbon of the
10 growing polyketide chain. A module can also contain domains for other enzymatic activities, such as, for example, a methyltransferase or dimethyltransferase activity. Finally, the releasing domain contains a thioesterase and, often, a cyclase activity.

The AT domain of the loading module recognizes a particular acyl-CoA (usually acetyl or propionyl but sometimes butyryl) and transfers it as a thiol ester to the ACP of the
15 loading module. Concurrently, the AT on each of the extender modules recognizes a particular extender-CoA (malonyl or alpha-substituted malonyl, i.e., methylmalonyl, ethylmalonyl, and carboxylglycolyl) and transfers it to the ACP of that module to form a thioester. Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module migrates to form a thiol ester (trans-esterification) at the KS of the first
20 extender module; at this stage, extender module 1 possesses an acyl-KS adjacent to a malonyl (or substituted malonyl) ACP. The acyl group derived from the loading module is then covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading unit (elongation or extension). The growing polyketide
25 chain is transferred from the ACP to the KS of the next module, and the process continues.

The polyketide chain, growing by two carbons each module, is sequentially passed as covalently bound thiol esters from module to module, in an assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the name polyketide arises. Most commonly,
30 however, additional enzymatic activities modify the beta keto group of each two-carbon unit just after it has been added to the growing polyketide chain, but before it is transferred to the next module. Thus, in addition to the minimal module containing KS, AT, and ACP domains necessary to form the carbon-carbon bond, modules may contain a ketoreductase (KR) that

reduces the keto group to an alcohol. Modules may also contain a KR plus a dehydratase (DH) that dehydrates the alcohol to a double bond. Modules may also contain a KR, a DH, and an enoylreductase (ER) that converts the double bond to a saturated single bond using the beta carbon as a methylene function. As noted above, modules may contain additional enzymatic activities as well.

Once a polyketide chain traverses the final extender module of a PKS, it encounters the releasing domain or thioesterase found at the carboxyl end of most PKSs. Here, the polyketide is cleaved from the enzyme and cyclized. The resulting polyketide can be modified further by tailoring enzymes; these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule.

While the above description applies generally to modular PKS enzymes, there are a number of variations that exist in nature. For example, some polyketides, such as epothilone, incorporate a building block that is derived from an amino acid. PKS enzymes for such polyketides include an activity that functions as an amino acid ligase or as a non-ribosomal peptide synthetase (NRPS). Another example of a variation, which is actually found more often than the two domain loading module construct found in DEBS, occurs when the loading module of the PKS is not composed of an AT and an ACP but instead utilizes an inactivated KS, an AT, and an ACP. This inactivated KS is in most instances called KS^Q, where the superscript letter is the abbreviation for the amino acid, glutamine, that is present instead of the active site cysteine required for activity. For example, the narbonolide PKS loading module contains a KS^Q. Yet another example of a variation has been mentioned above in the context of modules that include a methyltransferase or dimethyltransferase activity; modules can also include an epimerase activity. These variations will be described further below in specific reference to the narbonolide PKS and the various recombinant and hybrid PKSs provided by the invention.

With this general description of polyketide biosynthesis, one can better appreciate the biosynthesis of narbonolide related polyketides in *Streptomyces venezuelae* and *S. narbonensis*. The narbonolide PKS produces two polyketide products, narbonolide and 10-deoxymethynolide. Narbonolide is the polyketide product of all six extender modules of the narbonolide PKS. 10-deoxymethynolide is the polyketide product of only the first five extender modules of the narbonolide PKS. These two polyketides are desosaminylated to yield narbomycin and YC17, respectively. These two glycosylated polyketides are the final

The narbonolide PKS is composed of a loading module, six extender modules, and two thioesterase domains one of which is on a separate protein. Figure 4, part B, shows the organization of the narbonolide PKS genes on the *Streptomyces venezuelae* chromosome, as well as the location of the module encoding sequences in those genes, and the various domains within those modules. In the Figure, the loading module is not numbered, and its domains are indicated as sKS*, sAT, and ACP. Also shown in the Figure, part A, are the structures of picromycin and methymycin.

15 The loading and six extender modules and the thioesterase domain of the narbonolide
PKS reside on four proteins, designated PICA I, PICA II, PICA III, and PICA IV. PICA I
includes the loading module and extender modules 1 and 2 of the PKS. PICA II includes
extender modules 3 and 4. PICA III includes extender module 5. PICA IV includes extender
module 6 and a thioesterase domain. There is a second thioesterase domain (TE II) on a
20 separate protein, designated PICB. The amino acid sequences of these proteins are shown
below.

25	1	MSTVSKSESE	EFVSVSNDAG	SAHGTAEPVA	VVGISCRVPG	ARDPREFWEL	LAAGGQAVTD
	61	VPADRWNAGD	FYDPDRSAPG	RSNSRWGGFI	EDVDRFDAAF	FGISPREEAE	MDPQQRLALE
	121	LGWEALERAG	IDPSSLTGTR	TGVFAGAIWD	DYATLKHRQG	GAAITPHTVT	GLHRGIIANR
	181	LSYTLGLRGP	SMVVDSDGSS	SLVAVHLACE	SLRRGESELA	LAGGVSLNLV	PDSIIGASKF
30	241	GGLSPDGRAY	TFDARANGYV	RGEGGGFVVL	KRLSRAVADG	DPVLAVIRGS	AVNNGGAAQG
	301	MTTPDAQAQE	AVLREAHERA	GTAPADVRYV	ELHGTGTPVG	DPIEAAALGA	ALGTGRPAGQ
	361	PLLVGSVKTN	IGHLEGAAGI	AGLIKAVLAV	RGRALPASLN	YETPNPAIPF	EELNLRVNTE
	421	YLPWEPEHDG	QRMVVGVSSE	GMGGTNAHVV	LEEAPGVVEG	ASVVESTVGG	SAVGGGVVPW
35	481	VVSAKSAAAL	DAQIERLAAF	ASRDRTDGVD	AGAVDAGAVD	AGAVARVLAV	GRAQFEHRAV
	541	VVGSGPDDLA	AALAAPEGLV	RGVASGVGRV	AFVFPQGQTQ	WAGMGAELLD	SSAVFAAAMA
	601	ECEAALSPYV	DWSLEAVVRQ	APGAPTLERV	DVVQPVTFVAV	MVSLARVWQH	HGVTPQAVVG
	661	HSQGEIAAAY	VAGALSDDA	ARVVTLSRSK	IAAHLAGKGG	MLSLALSEDA	VLERLAGFDG
40	721	LSVAAVNGPT	ATVVSGDPVQ	IEELARACEA	DGVRARVIPV	DYASHSRQVE	IIESELAEVL
	781	AGLSPQAPRV	PFFSTLEGAW	ITEPVLDGGY	WYRNLRRHVG	FAPAVETLAT	DEGFTHFVEV
	841	SAHPVLTMAL	PGTVTGLATL	RRDNGGQDRL	VASLAEAWAN	GLAVDWSPLL	PSATGHHSOL
	901	PTYAFQTERH	WLGEIEALAP	AGEPAVQPAV	LRTEAAEPAE	LDRDEQLRVI	LDKVRAQTAQ
40	961	VLGYATGGQI	EVDRTFREAG	CTSLTGVDLR	NRINAAFGVR	MAPSMIFDFP	TPEALAEQLL

1021 LVVHGEAAAN PAGAEPAPVA AAGAVDEPVA IVGMACRLPG GVASPEDLWR LVAGGGDAIS
1081 EFPQDRGWDV EGLYHPDPEH PGTSYVRQGG FIENVAGFDA AFFGISPREA LAMDPQQRLL
1141 LETSWEAVED AGIDPDSLRLG RQVGVFTGAM THEYGPSLRD GGEGLDGYLL TGNTASVMSG
1201 RVSYTLGLEG PALTVDTACS SSLVALHLAV QALRKGEVDM ALAGGVAVMP TPGMFVEFSR
5 1261 QRGLAGDGRS KAFAASADGT SWSEGVGVLL VERLSDARRN GHQVLAVVRG SAVNQDGASN
1321 GLTAPNGPSQ QRVIRRALAD ARLTTSDDVDV VEAHGTGTRL GDPIEAQALI ATYGQGRDDE
1381 QPLRLGSLKS NIGHTQAAAG VSGVIKMQA MRHGLLPKTL HVDEPSDQID WSAGAVELLT
1441 EAVDWPEKQD GGLRRAAVSS FGISGTNAHV VLEEAPVVVE GASVVEPSVG GSAVGGGVTP
1501 WVVSASASAA LDAQIERLAA FASRDRTDDA DAGAVDAGAV AHVLADGRAQ FEHRAVALGA
10 1561 GADDLVQALA DPDGLIRGTA SGVGRVAFVF PGQGTQWAGM GAELLDSSAV FAAAMAECEA
1621 ALSPYVDWSL EAVVRQAPGA PTLERVDVVQ PVTFAVMVSL ARVWQHVGVT PQAVVGHSSQG
1681 EIAAAYVAGA LPLDDAARVV TLRKSIAAH LAGKGGMLSL ALNEDAVLER LSDFDGLSVA
1741 AVNGPTATVV SGDPVQIEEL AQACKADGFR ARIIPVDYAS HSRQVEIIES ELAQVLAGLS
1801 PQAPRVFFFS TLEGTWITEP VLDGTYWYRN LRHRVGFAPA IETLAVDEGF THFVEVSAHP
15 1861 VLTMTLPETV TGLGTLRREQ GGQERLVTSL AEAWVNGLPV AWTSLLPATA SRPGLPTYAF
1921 QAERYWLENT PAALATGDDW RYRIDWKRLP AAEGSERTGL SGRWLAVTPE DHSQAQAAVL
1981 TALVDAGAKV EVLTAGADDD REALAARLTA LTTGDGFTGV VSLLDGLVPQ VAWVQALGDA
2041 GIKAPLWSVT QGAVSVGRLD TPADPDRAML WGLGRVVALE HPERWAGLVD LPAQPDAAAL
2101 AHLVTALSGA TGEDQIAIRT TGLHARRLAR APLHGRRPTR DWQPHGTVLI TGGTGALGSH
20 2161 AARWMAHHGA EHLLVSRSG EQAPGATQLT AELTASGARV TIAACDVADP HAMRTLLDAI
2221 PAETPLTAVV HTAGALDDGI VDTLTAEQVR RAHRKAVGA SVLDELTRDL DLDAFVLFSS
2281 VSSTLGIPGQ GNYAPHNAYL DALAARRRAT GRSVSVVAVG PWDGGGMAAG DGVAERLRNH
2341 GVPGMDPELA LAALESALGR DETAITVADI DWDRFYLAIS SGRPQPLVEE LPEVRRIDA
2401 RDSATSGQGG SSAQGANPLA ERLAAAAPGE RTEILLGLVR AQAAAVLRMR SPEDVAADRA
25 2461 FKDIGFDSL AVELNRRLTR ATGLQLPATL VFDHPTPLAL VSLLRSEFLG DEETADARRS
2521 AALPATVGAG AGAGAGTDAD DDPIAIVAMS CRYPGDIRSP EDLWRMLSEG GEGITPFPTD
2581 RGWDLGLYD ADPDALGRAY VREGGFLHDA AEFDAEFFGV SPREALAMPD QORMLLTTSW
2641 EAFERAGIEP ASLRGSSTGV FIGLSYQDYA ARVPNAPRGV EGYLLTGSTP SVASGRIAYT
2701 FGLEGPAITV DTACSSSLTA LHLAVRALRS GECTMALAGG VAMMATPHMF VEFQRQALA
30 2761 PDGRSKAFSA DADGFGAAEG VGLLLVERLS DARRNGHPVL AVVRGTAVNQ DGASNGLTAP
2821 NGPSQQRVIR QALADARLAP GDIDAVETHG TGTSLGDPLE AQGLQATYCK ERPAERPLAI
2881 GSVKSNIGHT QAAAGAAGII KMLAMRHGT LPKTLHADEP SPHVDWANSK LALVTEPIDW
2941 PAGTGPRRAA VSSFGISGTN AHVVLEQAPD AGEVLGADE VPEVSETVAM AGTAGTSEVA
3001 EGSEASEAPA APGSREASLP GHLPPVLSAK DEQSLRGQAA ALHAWLSEPA ADLSDADGPA
35 3061 RLRDVGYTLA TSRTAFAHRA AVTAADRDFG LDGLATLAQG GTSAHVHLDL ARDGTAFLE
3121 TGQGSQRPGA GRELYDRHPV FARALDEICA HLDGHLELPL LDVMFAAEGS AEAALLDETR
3181 YTQCALFALE VALFRLVESW GMRPAALLGH SVGEIAAAHV AGVFSLADAA RLVAARGRLM
3241 QELPAGGAML AVQAAEDEIR VWLETEERYA GRLDVAAVNG PEAVALSGDA DAAREAEAYW
3301 SGLGRRTRAL RVSHAFHSAH MDGMLDGFRV VLETVEFRRP SLTVVSNVTG LAAGPDDLCD
40 3361 PEYWRHVRG TVRFLDGVRV LRDGLVTRCL ELGPDGVLTA MAADGLADTP ADSAAGSPVG
3421 SPAGSPADSA AGALRPRPLL VALLRRKRSE TETVADALGR AHAHGTGPDW HAWFAGSGAH
3481 RVDLPTYSFR RDRYWLDAFA ADTAVDTAGL GLGTADHPLL GAVVSLPDRD GLLLTGRLSL
3541 RTHPWLADHA VLGSVLLPGA AMVELAAHAA ESAGLRDVRE LTLLEPLVLP EHGGVELRVT
3601 VGAPAGEPGG ESAGDGARPV SLHSRLADAP AGTAWSCHAT GLLATDRPEL PVAPDRAAMW
45 3661 PPQGAEEVPL DGLYERLDGN GLAFGPLFQG LNAVWRYEGE VFADIALPAT TNATAPATAN
3721 GGSAAAAPY GIHPALLDAS LHAIAVGGVLV DEPELVVRPF HWSGTVVHAA GAAAARVRLA
3781 SAGTDAVSLS LTDGEGRPLV SVERLTLPV TADQAAASRV GGLMHRVAWR PYALASSGEQ
3841 DPHATSYGPT AVLKDELKV AALESAGVE VGLYPLDAL SQDVAAGAPA PRTVLAPLPA
3901 GPADGGAEGV RGTVARTLEL LQAWLADEHL AGTRLLLVTR GAVRDPEGSG ADDGGEDLSH
50 3961 AAAGLVRTA QTENPGRFGL LDLADASSY RTLPSVLSA GLRDEPQLAL HDGTIRLRL
4021 ASVRPETGTA APALAPEGTV LLTGGTGGLG GLVARHVVEG WGVRRLLLVV RRGTDAPGAD
4081 ELVHELEALG ADVSVAACDV ADREALTAVL DAIPAETHPL AVVHTAGVLS DGTLPSTMTTE
4141 DVEHVLKPKV DAAFLDELTA STPAYDLAAF VMFSSAAAVF GGAGQGAYAA ANATLDALAW
4201 RRAAGLPAL SLGWGLWAET SGMGTGELGQA DLRRMSRAGI GGISDAEGIA LLDAALRDDR
55 4261 HPVLLPLRLD AAGLRDAAGN DPAGIPALFR DVVGARTVRA RPSAASASTT AGTAGTPGTA
4321 DGAAETAAVT LADRAATVDG PARQRLLEF VVGEVAEVLG HARGHRIDAE RGFLDLGDFS
4381 LTAVELRNRL NSAGGLALPA TLVFDHPSA ALASHLDAEL PRGASDQDGA GNRNGNENGT
4441 TASRSTAETD ALLAQLTRLE GALVLTGLSD APGSEEVLEH LRSLSMVTG ETGTGTASGA
4501 PDGAGSGAED RPAAGDGAG GSEDGAGVP DFMNASAEEL FGLLDQDPST D (SEQ ID NO:1)

Amino acid sequence of narbonolide synthase subunit 2, PICAH (SEQ ID NO:2)

	1	VSTVNEEKYL	DYLRRATADL	HEARGRLREL	EAKAGEPVAI	VGMACRLPGG	VASPEDLWRL
	61	VAGGEDAISE	FPQDRGWDVE	GLYDPNPEAT	GKSYAREAGF	LYEAGEFDAD	FFGISPREAL
5	121	AMDPQQRLLL	EASWEAFEHA	GIPAATARGT	SVGVFTGVMY	HOYATRLTDV	PEGIEGYLGT
	181	GNSGSVASGR	VAYTLGLEGP	AVTVDTACSS	SLVALHLAVQ	ALRKGEVDMA	LAGGVTVMST
	241	PSTFVEFSRQ	RGLAPDGRSK	SFSSTADGTS	WSEGVGVLLV	ERLSDARRKG	HRILAVVRGT
	301	AVNQDGASSG	LTAPNGPSQQ	RVIRRALADA	RLTTSDDVV	EAHGTGTRLG	DPIEAQAVIA
	361	TYGQGRDGEQ	PLRLGSLKSN	IGHTQAAAGV	SGVIKMVQAM	RHGVLPKTLH	VEKPTDQVDW
10	421	SAGAVELLTE	AMDWPKGDG	GLRRAAVSSF	GVSGTNAHV	LEEAPAAEET	PASEATPAVE
	481	PSVGAGLVPW	LVS AKTPAAL	DAQIGRLAAF	ASQGRDAAD	PGAVARVLG	GRAEFEHRAV
	541	VLGTGQDDFA	QALTAPEGLI	RGTPSDVGRV	AFVFPQGQTQ	WAGMGAELLD	VSKEFAAAMA
	601	ECESALSRYV	DWSLEAVVRQ	APGAPTLERV	DVVQPVTFV	MVSLAKVWQH	HGVTPOAVVG
	661	HSQGEIAAAY	VAGALTDDA	ARVVTLSKS	IAAHLAKGG	MISLALSEEA	TRORIENLHG
15	721	LSIAAVNGPT	ATVVS GDPTQ	IQELAQACEA	DGVRARIIPV	DYASHSAHVE	TIESELAEVL
	781	AGLSPRTPEV	PFFSTLEGAW	ITEPVLDTY	WYRNLRRHVG	FAPAVETLAT	DEGFTHFIEV
	841	SAHPVLTMTL	PETVTGLGTL	RREQGGQERL	VTSLAEAWTN	GLTIDWAPVL	PTATGHHPEL
	901	PTYAFQRRHY	WLHDS PAVQG	SVQDSWRYRI	DWKRLAVADA	SERAGLSGRW	LVVVPEDRSA
	961	EAAPVLAALS	GAGADPVQLD	VSPLGDRQRL	AATLGEALAA	AGGAVDGVLS	LLAWDESAHP
20	1021	GHPAPFTRGT	GATLTLVQAL	EDAGVAAPLW	CVTHGAVSVG	RADHVTSPAQ	AMVWGMGRVA
	1081	ALEHPERWGG	LIDLPSDADR	AALDRMTTVL	AGGTGEDQVA	VRASGLLARR	LVRASLPAHG
	1141	TASPWWQADG	TVLVTGAEEP	AAAEAAARLA	RDGAGHLLH	TPSGSEGAE	GTSGAAEDSG
	1201	LAGLVAELAD	LGATATVVT	DLTDAEAAAR	LLAGVSDAHP	LSAVLHLPPT	VDSEPLAATD
	1261	ADALARVVTA	KATAALHLDR	LLREAAAAGG	RPPVLVLFSS	VAAIWGGAGQ	GAYAAGTAF
25	1321	DALAGQHRAD	GPTVTSVAWS	PWEGSRVTEG	ATGERLRLRG	LRPLAPATAL	TALDTALGHG
	1381	DTAVTIADVD	WSSFAPGFTT	ARPGTLLADL	PEARALDEQ	QSTTAADDTV	LSRELGALTG
	1441	AEQQRMQEL	VREHLAVVLN	HPSPEAVDTG	RAFRDLGFDS	LTAVELRNRL	KNATGLALPA
	1501	TLVFDYPTPR	TLAEFLLAEI	LGEQAGAGEQ	LPVDGGVDDE	PVAIVGMACR	LPGGVASPED
	1561	LWRLVAGGED	AISGFPQDRG	WDVEGLYDPD	PDASGRTYCR	AGGFLDEAGE	FDADFFGISP
30	1621	REALAMDPQQ	RLLLET SWEA	VEDAGIDPTS	LQGOQGVVFA	GTNGPHYEPL	LRNTAEDLEG
	1681	YVGTGNAASI	MSGRVSYTLG	LEGPAVTVD	ACSSSLVALH	LAVQALRKGE	CGLALAGGVT
	1741	VMSTPTTFVE	FSRQRLAED	GRSKAFAASA	DGFGPAEGVG	MLLVERLSDA	RRNGHRVLAV
	1801	VRGSAVNQDG	ASNGLTAPNG	PSQQRVIRRA	LADARLTAD	VDVVEAHGTG	TRLGDPIEAQ
	1861	ALIATYGQGR	DTEQPLRLGS	LKSNIGHTQA	AAGVSGIIM	VQAMRHGVLP	KTLHVDPRSD
35	1921	QIDWSAGTVE	LLTEAMDWPR	KQEGGLRRAA	VSSFGISGTN	AHIVLEEAPV	DEDAPADEPS
	1981	VGGVVPWLVS	AKTPAALDAQ	IGRLAAFASQ	GRTDAADPGA	VARVLAGGRA	QFEHRAVALG
	2041	TGQDDLAAAL	AAPEGLVRGV	ASGVGRVAFV	FPGQGTQWAG	MGAELLDVSK	EFAAAMAECE
	2101	AALAPYVDWS	LEAVVRQAPG	APTLEVRDVV	QPVTFVAVMS	LAKVWQHGV	TPQAVVGHSSQ
	2161	GEIAAAYVAG	ALSLDDAARV	VTLSKSSIGA	HLAQGGMLS	LALSEA AVE	RLAGFDGLSV
40	2221	AAVNGPTATV	VSGDPTQIQE	LAQACEADGV	RARIIPVDYA	SHSAHVETIE	SELADVLAGL
	2281	SPQTPQVPFF	STLEGAWITE	PALDGGYWYR	NLRHRVGFAP	AVETLATDEG	FTHFVEVSAH
	2341	PVLTMALPET	VTGLGTLRRD	NGGQHRLTTS	LAEAWANGLT	VDWASLLPTT	TTHPDLPTYA
	2401	FQTERYWPQP	DLAAGDITS	AGLGAAEHPL	LGAVALADS	DGCLLTGSL	LRTHPWLADH
	2461	AVAGTVLLPG	TAFVELAFRA	GDQVGC DLVE	ELTLDAPLVL	PRRGAVRVQL	SVGASDESGR
45	2521	RTFGLYAHPE	DAPGEAEWTR	HATGVLAARA	DRTAPVADPE	AWPPPGAEPV	DVDGLYERFA
	2581	ANGYGYGPLF	QGVRGVWRRG	DEVFADVALP	AEVAGAEGAR	FGLHPALLDA	AVQAAGAGGA
	2641	FGAGTRL PFA	WSGISLYAVG	ATALRVRLAP	AGPDTVS VSA	ADSSGQPVFA	ADSLTVLPVD
	2701	PAQLAAFSDP	TLDALHLEW	TAWDGAAQAL	PGAVVLGGDA	DGLAAALRAG	GTEVLSFPDL
	2761	TDLVEAVDRG	ETPAPATVLV	ACPAAGPGGP	EHVREALHGS	LALMQAWLAD	ERFTDGRVL
50	2821	VTRDAVAARS	GDGLRSTGQA	AVWGLGRSAQ	TESPGRFVLL	DLAGEARTAG	DATAGDGLTT
	2881	GDATEVGGTSG	DAALGSALAT	ALGSSEPQLA	LRDGALLVPR	LARAAAPAAA	DGLAAADGLA
	2941	ALPLPAAPAL	WRLEPGTDGS	LESLTAAPGD	AETLAPEPLG	PGQVRIAIRA	TGLNFRDVLI
	3001	ALGMYPD PAL	MGTEGAGVVT	ATGPGVTHLA	PGDRVMGLLS	GAYAPVVVAD	ARTVARMPEG
	3061	WTFAQGASVP	VVFLTAVYAL	RDLADV KPG	RLLVHSAAGG	VGMAAVQLAR	HWGVEVHGTA
55	3121	SHGKWDALRA	LGLDDAHIAS	SRTLDFESAF	RAASGGAGMD	VVLNSLAREF	VDASLRL LGP
	3181	GGRFVEMGKT	DVRDAERVAA	DHPGVGYRAF	DLGEAGPERI	GEMLAEVIAL	FEDGVLRHLP
	3241	VTTWDVRRAR	DAFRHVSQAR	HTGKVLTMP	SGLDPEGTVL	LTGGTGALGG	IVARHVVG EW
	3301	GVRRLLLVSR	RGTDPAGAGE	LVHELEALGA	DVSVAACDVA	DREALTAVLD	SIPA EHPLTA

3361 VVHTAGVLSD GTLPSMTAED VEHVLRPKVD AAFLLDELTS TPGYDLAAV MFSSAAAVFG
 3421 GAGQGAYAAA NATLDALAWR RRTAGLPALS LGWGLWAETS GMTGGLSDTD RSRLARSGAT
 3481 PMDSELTLSL LDAAMRRDDP ALVPIALDVA ALRAQQRDGM LAPLLSGLTR GSRVGGAPVN
 3541 QRRAAAGGAG EADTDLGGR LRLPATLVFD HPTPGELAGH LLDELATAAG GSWAEGTGSG
 5 3601 TGFDSLTAVE LRNRLNAATG LRLPATLVFD HPTPGELAGH LLDELATAAG GSWAEGTGSG
 3661 DTASATDRQT TAALAEELDRL EGVLASLAPA AGGRPELAAR LRALAAALGD DGDDATDLDE
 3721 ASDDDLFSFI DKELGDSDF (SEQ ID NO:2)

Amino acid sequence of narbonolide synthase subunit 3, PICAIII (SEQ ID NO:3)

10 1 MANNEDKLRD YLKRVTAELQ QNTRRLREIE GRTHPEVAIV GMACRLPGGV ASPEDLWQLV
 61 AGDGDASEF PQDRGWDVEG LYDPDPDASG RTYCRSGGFL HDAGEFDADF FGISPREALA
 121 MDPQQRSLT TAWAIESAG IDPTALKGSG LGVFVGGWHT GYTSGOTTAV QSPELEGLHV
 181 SGAALGFLSG RIAYVLGTDG PALTVDTACS SSLVALHLAV QALRKGECDM ALAGGVTVMP
 241 NADLFVQFSR QRGLAADGRS KAFATSADGF GPAEGAGVLL VERLSDARRN GHRILAVVRG
 15 301 SAVNQDGASN GLTAPHGPSQ QRVIRRALAD ARLAPGDVDV VEAHGTGTRL GDPIEAQALI
 361 ATYGQEKSS QPLRLGALKS NIGHTQAAAG VAGVIKMQA MRHGLLPKTL HVDEPSDQID
 421 WSAGTVELLT EAVDWPEKQD GGLRRAAVSS FGISGTNAHV VLEEAPAVED SPAVEPPAGG
 481 GVPWPVSAK TPAALDAQIG QLAAYADGRT DVDPAAVAAA LVDSRTAMEH RAVAVGDSRE
 541 ALRDALRMPE GLVRGTSSDV GRVAFVFPQG GTQWAGMGAE LLDSSPEFAA SMAECETALS
 20 601 RYVDWSLEAV VRQEPGAPTL DRVDVVPQT FAVMVSLAKV WQHHGITPQA VVGHSQGEIA
 661 AAYVAGALT DDAARVVTLR SKSIAAHLAG KGGMISLALD EAAVLKRLSD FDGLSVAASN
 721 GPTATVSGD PTQIEELART CEADGVRARI IPVDYASHSR QVEIIEKELA EVLAGLAPQA
 781 PHVPFFSTLE GTWITEPVLG GTYWYRNLRH RVGFAPAVET LAVDGFTHFI EVSAHPVLTM
 841 TLPETVTGLG TLRREQGGQE RLVTSLAEAW ANGLTIDWAP ILPTATGHHP ELPTYAFQTE
 25 901 RFWLQSSAPT SAADDWRYRV EWKPLTASGQ ADLSGRWIVA VGSEPEAELL GALKAGAEEV
 961 DVLEAGADDD REALAARLT LTTGDGFTGV VSLDDLVPQ VAWVQALGDA GIKAPLWSVT
 1021 QGAVSVGRDL TPADPDRAML WGLGRVVALE HPERWAGLVD LPAQPDAAAL AHLVTALSGA
 1081 TGEDQIAIRT TGLHARRLAR APLHGRRPTR DWQPHGTVLI TGGTGALGSH AARWMAHGA
 1141 EHLLLVSRSG EQAPGATQLT AELTASGARV TIAACDVADP HAMRTLDAI PAETPLTAVV
 30 1201 HTAGAPGGDP LDVTGPEDIA RILGAKTSGA EVLDDLRTG PLDAFVLYSS NAGVWGSQSG
 1261 GYAAANNAHL DALAARRRAR GETATSVAWG LWAGDGMGRG ADDAYWQRRG IRPMSPRAL
 1321 DELAKALSHD ETFVAVADV WERFAPAFV SRPSLLLDGV PEARQALAAP VGAPAPGDAA
 1381 VAPTQSSAL AAITALPEPE RRPALLTLVR THAAVLGHS SPDRVAPGRA FTELGFDSL
 1441 AVQLRNQLST VVGNRLPATT VFDHPTPAL AAHLHEAYLA PAEPAPTDWE GRVRRALAE
 35 1501 PLDRLRDAGV LDTVLRGTI EPEPGSGSD GGAADPGAEP EASIDDLDAE ALIRMALGPR
 1561 (SEQ ID NO:3)

Amino acid sequence of narbonolide synthase subunit 4, PICAIV (SEQ ID NO:4)

40 1 MTSSNEQLVD ALRASLKENE ELRKESRRRA DRRQEPMAIV GMSCRFAGGI RSPEDLWDAV
 61 AAGKDLVSEV PEERGWDIDS LYDPVPGRKG TTYVRNAFL DDAAGFDAF FGISPREALA
 121 MDPQQRQLE ASWEVFERAG IDPASVRGTD VGVVVGCGYQ DYAPDIRVAP EGTGGYVVTG
 181 NSSAVASGRI AYSLGLEGPA VTVDTACSSS LVALHLALKG LRNGDCSTAL VGGVAVLATP
 241 GAFIEFSSQQ AMAADGRTKG FASAADGLAW GEGVAVLLE RLSDARRKGH RVLAVVRGSA
 301 INQDGASNG L TAPHGPSQQR LIRQALADAR LTSSDQDVVE GHGTGTRLGD PIEAQALLAT
 45 361 YGQGRAPGQP LRLGTLKSN I GHTQAASGVA GVIKMQALR HGVLPKTLHV DEPTDQVDWS
 421 AGSVELLTEA VDWPFRPGR L RAGVSAFGV GGTNAHVVE EAPAVEESPA VEPPAGGGV
 481 PWPVSAKTS ALDAQIGQLA AYAEDRTDV PAVAAALVD SRTAMEHRAV AVGDSREALR
 541 DALRMPEGLV RGTVDTPGRV AFVFPQGQTQ WAGMGAE LLD SSPEFAAAMA ECETALSPYV
 601 DWSLEAVVRQ APSAPTLDRV DVVQPVTFV MVSLAKVWQH HGITPEAVIG HSQGEIAAAY
 50 661 VAGALTDDA ARVVTLRSKS IAAHLAKGG MISLALSEE TRQRIENLHG LSIAAVNGPT
 721 ATVVSGDPTQ IQELAQACEA DGIRARIIPV DYASHSAHVE TIENELADV AGLSPTPQV
 781 PFFSTLEGTW ITEPALDGGY WYRNLRRHV FAPAVETLAT DEGFTHFIEV SAHPVLTMTL
 841 PDKVTGLATL RREDGGQHRL TTSLAEAWAN GLALDWASLL PATGALSPAV PDLPTYAFQH
 901 RSYWISPAGP GEAPANTASG REAVAETGLA WPGAEDLDE EGRRSAVLAM VMRQAASVLR
 55 961 CDSPEEVPD RPLREIGFDS LTAVDFNRV NRTGLQLPP TVVFEHPTPV ALAERISDEL
 1021 AERNWAVEP SDHEQAEEEK AAPAGARSG ADTGAGAGMF RALFRQAVED DRYGEFLDVL

1081 AEASAFRPQF ASPEACSERL DPVLLAGGPT DRAEGRAVLV GCTGTAANGG PHEFLRLSTS
 1141 FQEERDFLAV PLPGYGTGTG TGTALLPADL DTALDAQARA ILRAAGDAPV VLLGHSGGAL
 1201 LAHELAFRLE RAHGAPPAGI VLVDPPYPGH QEPIEVWSRQ LGEGLFAGEL EPMSDARLLA
 1261 MGRIYARFLAG PRPGRSSAPV LLVRASEPLG DWQEERGDWR AHWDLPH TVA DVPGDHFTMM
 5 1321 RDHAPAVAEA VLSWLDAIEG IEGAGK (SEQ ID NO:4)

Amino acid sequence of typeII thioesterase, PICB (SEQ ID NO:5)

1 VTDRPLNVDS GLWIRRFHPA PNSAVRLVCL PHAGGSASYF FRFSEELHPS VEALSVQYPG
 61 RQDRRAEPCL ESVEELAEHV VAATEPWWQE GRLAFFGHSL GASVAFETAR ILEQRHGVRP
 10 121 EGLYVSGRRA PSLAPDRLVH QLDDRAFLAE IRRLSGTDER FLQDDELLRL VLPALRS DYK
 181 AAETYLHRPS AKLTCPVMAL AGDRDPKAPL NEVAEWRRT SGPFCL RAYS GGHFY LNDQW
 241 HEICNDISDH LLVTRGAPDA RVVQPPTS LI EGA AKRWQNP R (SEQ ID NO:5)

The DNA encoding the above proteins can be isolated in recombinant form from the
 15 recombinant cosmid pKOS023-27 of the invention, which was deposited with the American
 Type Culture Collection under the terms of the Budapest Treaty on 20 August 1998 and is
 available under accession number ATCC 203141. Cosmid pKOS023-27 contains an insert of
Streptomyces venezuelae DNA of ~38506 nucleotides. The complete sequence of the insert
 from cosmid pKOS023-27 is shown below. The location of the various ORFs in the insert, as
 20 well as the boundaries of the sequences that encode the various domains of the multiple
 modules of the PKS, are summarized in the Table below. Figure 2 shows a restriction site and
 function map of pKOS023-27, which contains the complete coding sequence for the four
 proteins that constitute narbonolide PKS and four additional ORFs. One of these additional
 ORFs encodes the *picB* gene product, the type II thioesterase mentioned above. PICB shows
 25 a high degree of similarity to other type II thioesterases, with an identity of 51%, 49%, 45%
 and 40% as compared to those of *Ammycolatopsis mediterranea*, *S. griseus*, *S. fradiae* and
Saccharopolyspora erythraea, respectively. The three additional ORFs in the cosmid
 pKOS023-27 insert DNA sequence, from the *picCII*, *picCIII*, and *picCVI*, genes, are involved
 in desosamine biosynthesis and transfer and described in the following section.

30

	From Nucleotide	To Nucleotide	Description
	70	13725	<i>picAI</i>
	70	13725	narbonolide synthase 1 (PICA1)
	148	3141	loading module
35	148	1434	KS loading module
	1780	2802	AT loading module
	2869	3141	ACP loading module
	3208	7593	extender module 1
	3208	4497	KS1
40	4828	5847	AT1

	6499	7257	KR ¹
	7336	7593	ACP1
	7693	13332	extender module 2
	7693	8974	KS2
5	9418	10554	AT2
	10594	11160	DH2
	12175	12960	KR2
	13063	13332	ACP2
	13830	25049	<i>picAII</i>
10	13830	25049	narbonolide synthase 2 (PICAII)
	13935	18392	extender module 3
	13935	15224	KS3
	15540	16562	AT3
	17271	18071	KR3 (inactive)
15	18123	18392	ACP3
	18447	24767	extender module 4
	18447	19736	KS4
	20031	21050	AT4
	21093	21626	DH4
20	22620	23588	ER4
	23652	24423	KR4
	24498	24765	ACP4
	25133	29821	<i>picAIII</i>
	25133	29821	narbonolide synthase 3 (PICAIII)
25	25235	29567	extender module 5
	25235	26530	KS5
	26822	27841	AT5
	28474	29227	KR5
	29302	29569	ACP5
30	29924	33964	<i>picAIV</i>
	29924	33964	narbonolide synthase 4 (PICAIV)
	30026	32986	extender module 6
	30026	31312	KS6
	31604	32635	AT6
35	32708	32986	ACP6
	From Nucleotide	To Nucleotide	Description
	33068	33961	PKS thioesterase domain
	33961	34806	<i>picB</i>
	33961	34806	typeII thioesterase homolog
40	34863	36011	<i>picCII</i>
	34863	36011	4-keto-6-deoxyglucose isomerase
	36159	37439	<i>picCIII</i>
	36159	37439	desosaminyl transferase
	37529	38242	<i>picCVI</i>
45	37529	38242	3-amino dimethyltransferase

DNA Sequence of the Insert DNA in Cosmid pKOS023-27 (SEQ ID NO:19)

1 GATCATGCGG AGCACTCCTT CTCTCGTGCT CCTACCGGTG ATGTGCGCGC CGAATTGATT

5 61 CGTGGAGAGA TGTCGACAGT GTCCAAGAGT GAGTCCGAGG AATTCGTGTC CGTGTGCAAC
121 GACGCCGGTT CCGCGCACGG CACAGCGGAA CCCGTCGCCG TCGTCGGCAT CTCCTGCCGG
181 GTGCCCCGGC CCGGGGACCC GAGAGAGTTC TGGGAAGTCC TGGCGGCAGG CGGCCAGGCC
241 GTCACCGACG TCCCCGCGGA CCGCTGGAAC GCCGGCGACT TCTACGACCC GGACCGCTCC
301 GCCCCCGGCC GCTCGAACAG CCGGTGGGGC GGGTTCATCG AGGACGTCGA CCGGTTCGAC
361 GCCGCCTTCT TCGGCATCTC GCCCCGCGAG GCCGCGGAGA TGGACCCGCA GCAGCGGCTC
421 GCCCTGGAGC TGGGCTGGGA GGCCCTGGAG CGCGCCGGGA TCGACCCGTC CTCGCTCACC
481 GGCACCCGCA CCGGCGTCTT CGCCGGCGCC ATCTGGGACG ACTACGCCAC CCTGAAGCAC
541 CGCCAGGGCG GCGCCGCGAT CACCCGCGAC ACCGTCACCG GCCTCCACCG CGGCATCATC
10 601 GCGAACCGAC TCTCGTACAC GCTCGGGCTC CGCGGCCCA GCATGGTCGT CGACTCCGGC
661 CAGTCCTCGT CGCTCGTCGC CGTCCACCTC GCGTGCGAGA GCCTGCGGCG CGGCGAGTCC
721 GAGCTCGCCC TCGCCGGCGG CGTCTCGCTC AACCTGGTGC CGGACAGCAT CATCGGGGCG
781 AGCAAGTTCG GCGGCCTCTC CCCCAGCGGC CGCGCCTACA CCTTCGACGC GCGCGCAAC
841 GGCTACGTAC GCGGCGAGGG CGGCGGTTTC GTCGTCTGA AGCGCCTCTC CCGGGCCGTC
15 901 GCCGACGGCG ACCCGGTGCT CGCCGTGATC CGGGGCAGCG CCGTCAACAA CGGCGGCGCC
961 GCCCAGGGCA TGACGACCCC CGACGCGCAG GCGCAGGAGG CCGTGCTCCG CGAGGCCAC
1021 GAGCGGGCCG GGACCGCGCC GGCCGACGTG CGGTACGTCG AGCTGCACGG CACCGGCACC
1081 CCCGTGGGCG ACCCGATCGA GGCCGCTGCG CTCGGCGCCG CCCTCGGCAC CGGCCGCCCG
1141 GCCGACAGC CGCTCCTGGT CGGCTCGGTC AAGACGAACA TCGGCCACCT GGAGGGCGCG
20 1201 GCCGGCATCG CCGGCCTCAT CAAGGCCGTC CTGGCGGTCC GCGGTCGCGC GCTGCCCGCC
1261 AGCCTGAACT ACGAGACCCC GAACCCGGCG ATCCCGTTTC AGGAACTGAA CCTCCGGGTG
1321 AACACGGAGT ACCTGCCGTG GGAGCCGGAG CACGACGGGC AGCGGATGGT CGTCGGCGTG
1381 TCCTCGTTTC GCATGGGCGG CACGAACGCG CATGTCTGTC TCGAAGAGGC CCCGGGGGTT
1441 GTCGAGGGTG CTTCCGGTCGT GGAGTCGACG GTCGGCGGGT CGGCGGTGCG CGGCGGTGTG
25 1501 GTGCCGTGGG TGGTGTGCGC GAAGTCCGCT GCCGCGCTGG ACGCGCAGAT CGAGCGGCTT
1561 GCCGCGTTTC CCTCGCGGGA TCGTACGGAT GGTGTGACG CGGGCGCTGT CGATGCGGGT
1621 GCTGTGATG CCGGTGCTGT CGCTCGCGTA CTGGCCGGCG GCGGTGCTCA GTTCGAGCAC
1681 CGGGCCGTCG TCGTCGGCAG CGGGCCGGAC GATCTGGCGG CAGCGCTGGC CGCGCCTGAG
1741 GGTCTGGTCC GGGGCGTGGC TTCCGGTGTC GGGCGAGTGG CGTTCGTGTT CCCCAGGCG
30 1801 GGCACGAGT GGGCCGGCAT GGGTGCCGAA CTGCTGGACT CTTCCGCGGT GTTCGCGGCG
1861 GCCATGGCCG AATGCGAGGC CGCACTCTCC CCGTACGTCG ACTGGTCGCT GGAGGCCGTC
1921 GTACGGCAGG CCCCCGGTGC GCCACGCTG GAGCGGGTCG ATGTCTGTGA GCCTGTGACG
1981 TTCGCCGTCA TGGTCTCGCT GGCTCGCGTG TGGCAGCACC ACGGGGTGAC GCCCCAGGCG
2041 GTCGTCGGCC ACTCGCAGGG CGAGATCGCC GCCGCGTACG TCGCCGGTGC CCTGAGCCTG
35 2101 GACGACGCCG CTCGTGTCTG GACCCTGCGC AGCAAGTCCA TCGCCGCCCA CCTCGCCGGC
2161 AAGGGCGGCA TGCTGTCCCT CGCGCTGAGC GAGGACGCCG TCCTGGAGCG ACTGGCCGGG
2221 TTCGACGGGC TGTCCGTGCG CGCTGTGAAC GGGCCACCG CCACCGTGGT CTCCGGTGAC
2281 CCCGTACAGA TCGAAGAGCT TGCTCGGGCG TGTGAGGCCG ATGGGGTCCG TGCGCGGGTC
2341 ATTCCCCTCG ACTACGCGTC CCACAGCCGG CAGGTGAGAG TCATCGAGAG CGAGCTCGCC
40 2401 GAGGTCTCTG CCGGGCTCAG CCCGAGGCT CCGCGCGTGC CGTTCCTCTC GAACTCGAA
2461 GGCCTGGA TCACCGAGCC CGTGCTCGAC GGCGGCTACT GGTACCGCAA CCTGCGCCAT
2521 CGTGTGGGCT TCGCCCCGGC CGTCGAGACC CTGGCCACCG ACGAGGGCTT CACCACTTC
2581 GTCGAGGTCA GCGCCACCC CGTCCTCACC ATGGCCCTCC CCGGGACCGT CACCGGTCTG
2641 GCGACCCTGC GTCGCGACAA CGGCGGTGAG GACCGCTCG TCGCCTCCCT CGCCGAAGCA
45 2701 TGGGCCAACG GACTCGCGGT CGACTGGAGC CCGCTCCTCC CCTCCGCGAC CGGCCACCAC
2761 TCCGACCTCC CCACCTACGC GTTCCAGACC GAGCGCCACT GGCTGGGCGA GATCGAGGCG
2821 CTCGCCCCGG CCGGCGAGCC GGCGGTGAG CCCGCCGTCC TCCGCACGGA GGCGGCCGAG
2881 CCGGCGGAGC TCGACCGGGA CGAGCAGCTG CGCGTGATCC TGGACAAGGT CCGGGCGCAG
2941 ACGGCCAGG TGCTGGGGTA CGCGACAGGC GGGCAGATCG AGGTGACCG GACCTTCCGT
50 3001 GAGGCCGGTT GCACCTCCCT GACCGGCGTG GACCTGCGCA ACCGGATCAA CGCCGCCTTC
3061 GCGGTACGGA TGGCGCCGTC CATGATCTTC GACTTCCCCA CCCCCAGGC TCTCGCGGAG
3121 CAGCTGCTCC TCGTCGTGCA CGGGGAGGCG GCGGCGAACC CGGCCGGTGC GGAGCCGGCT
3181 CCGGTGGCGG CCGCCGGTGC CGTCGACGAG CCGGTGGCGA TCGTCGGCAT GGCCTGCCGC
3241 CTGCCCCGGT GGGTGCCTC GCCGGAGGAC CTGTGGCGGC TGGTGGCCGG CGGCGGGGAC
55 3301 GCGATCTCGG AGTTCCCGCA GGACCGCGGC TGGGACGTGG AGGGGCTGTA CCACCCGGAT
3361 CCCGAGCACC CCGGCACGTC GTACGTCCGC CAGGGCGGTT TCATCGAGAA CGTCGCCGGC
3421 TTCGACGCGG CCTTCTTCGG GATCTCGCCG CGCGAGGCC TCGCCATGGA CCCGACGAG
3481 CGGCTCCTCC TCGAAACCTC CTGGGAGGCC GTCGAGGACG CCGGGATCGA CCGACCTCC
3541 CTGCGGGGAC GGCAGGTGCG CGTCTTCACT GGGGCGATGA CCCACGAGTA CGGGCCGAGC

3601 CTGCGGGACG GCGGGGAAGG CCTCGACGGC TACCTGCTGA CCGGCAACAC GGCCAGCGTG
3661 ATGTCGGGCC GCGTCTCGTA CACACTCGGC CTTGAGGGCC CCGCCCTGAC GGTGGACACG
3721 GCCTGCTCGT CGTCGCTGGT CGCCCTGCAC CTCGCCGTGC AGGCCCTGCG CAAGGGCGAG
3781 GTCGACATGG CGCTCGCCGG CGGCGTGGCC GTGATGCCCA CGCCCGGGAT GTTCGTGAG
5 3841 TTCAGCCGGC AGCGCGGGCT GGCCGGGGAC GGCCGGTCGA AGGCGTTCGC CGCGTCGGCG
3901 GACGGCACCA GCTGGTCCGA GGGCGTCGGC GTCCTCCTCG TCGAGCGCCT GTCGGACGCC
3961 CGCCGCAACG GACACCAGGT CCTCGCGGTC GTCCGCGGCA GCGCCGTGAA CCAGGACGGC
4021 GCGAGCAACG GCCTCACGGC TCCGAACGGG CCCTCGCAGC AGCGCGTCAT CCGGCGCGCG
4081 CTGGCGGACG CCCGGCTGAC GACCTCCGAC GTGGACGTCG TCGAGGCACA CGGCACGGGC
10 4141 ACGCGACTCG GCGACCCGAT CGAGGCGCAG GCCCTGATCG CCACCTACGG CCAGGGCCGT
4201 GACGACGAAC AGCCGCTGCG CCTCGGGTCG TTGAAGTCCA ACATCGGGCA CACCCAGGCC
4261 GCGGCCGGCG TCTCCGGTGT CATCAAGATG GTCCAGGCGA TCGGCCACGG ACTGCTGCCG
4321 AAGACGCTGC ACGTCGACGA GCCCTCGGAC CAGATCGACT GGTCGGCTGG CGCCGTGGAA
4381 CTCCTACCG AGGCCGTCGA CTGGCCGGAG AAGCAGGACG GCGGGCTGCG CCGGGCCGCC
15 4441 GTCTCCTCCT TCGGGATCAG CGGCACCAAT GCGCATGTGG TGCTCGAAGA GGCCCCGGTG
4501 GTTGTGAGG GTGCTTCGGT CGTCGAGCCG TCGGTGCGCG GGTGCGCGGT CCGCGGCGGT
4561 GTGACGCCTT GGGTGGTGTG GCGAAGTCC GCTGCCGCGC TCGACGCGCA GATCGAGCGG
4621 CTTGCCGCAT TCGCCTCGCG GGATCGTACG GATGACGCCG ACGCCGGTGC TGTCGACCG
4681 GCGGCTGTGCT CTCACGTACT GGCTGACGGG CGTGCTCAGT TCGAGCACCG GGCCGTGCGG
20 4741 CTCGGCGCCG GGGCGGACGA CCTCGTACAG GCGCTGGCCG ATCCGGACGG GCTGATACGC
4801 GGAACGGCTT CCGGTGTCGG GCGAGTGGCG TTCGTGTTCC CCGGTCAGGG CACGAGTGG
4861 GCTGGCATGG GTGCCGAAC TCTGGACTCT TCCGCGGTGT TCGCGCGGC CATGGCCGAG
4921 TGTGAGGCCG CGCTGTCCCC GTACGTCGAC TGGTCGCTGG AGGCCGTCTG ACGGCAGGCC
4981 CCCGGTGCGC CCACGCTGGA GCGGGTCGAT GTCGTGCAGC CTGTGACGTT CGCCGTCTATG
25 5041 GTCTCGCTGG CTCGCGTGTG GCAGACCAC GGTGTGACGC CCCAGGCGGT CGTCGGCCAC
5101 TCGCAGGGCG AGATCGCCGC CGCGTACGTC GCCGGAGCCC TGCCCCCTGGA CGACGCCGCC
5161 CGCGTCGTCA CCCTGCGCAG CAAGTCCATC GCCGCCACC TCGCCGCAA GGGCGGCATG
5221 CTGTCCCTCG CGCTGAACGA GGACGCCGTC CTGGAGCGAC TGAGTGAATT CGACGGGCTG
5281 TCCGTGCGCG CCGTCAACGG GCCACCGCC ACTGTCGTGT CGGGTGACCC CGTACAGATC
30 5341 GAAGAGCTTG CTCAGGCGTG CAAGGCGGAC GGATTCGCG CGCGGATCAT TCCCGTCGAC
5401 TACGCGTCCC ACAGCCGGCA GGTCGAGATC ATCGAGAGCG AGCTCGCCA GGTCTCGCC
5461 GGTCTCAGCC CGCAGGCCCC GCGCGTGCCG TTCTTCTCGA CGCTCGAAGG CACCTGGATC
5521 ACCGAGCCCG TCCTCGACGG CACCTACTGG TACCGCAACC TCCGTACCG CGTCGGCTTC
5581 GCCCCCGCCA TCGAGACCTT GGCCGTGAC GAGGGCTTCA CGCACTTCGT CGAGGTCAGC
35 5641 GCCCACCCCG TCCTACCAT GACCCTCCCC GAGACCGTCA CCGGCCTCGG CACCCTCCGT
5701 CGCGAACAGG GAGGCCAAGA GCGTCTGGTC ACCTCGCTCG CCGAGGCGTG GGTCAACGGG
5761 CTTCCCGTGG CATGGACTTC GCTCCTGCCC GCCACGGCCT CCCGCCCCGG TCTGCCACC
5821 TACGCCTTCC AGGCCGAGCG CTACTGGCTC GAGAACAATC CCGCCGCCCT GGCCACCGGC
5881 GACGACTGGC GCTACCGCAT CGACTGGAAG CGCTCCCGG CCGCCGAGGG GTCCGAGCGC
40 5941 ACCGGCCTGT CCGGCCGCTG GTCGCGCTC ACGCCGAGG ACCACTCCGC GCAGGCCGCC
6001 GCCGTGCTCA CCGCGCTGGT CGACGCCGGG GCGAAGGTCG AGGTGCTGAC GGCCGGGGCG
6061 GACGACGACC GTGAGGCCCT CGCCGCCCGG CTCACCGCAC TGACGACCGG TGACGGCTTC
6121 ACCGGCGTGG TCTCGCTCCT CGACGGACTC GTACCGCAGG TCGCCTGGGT CCAGGCGCTC
6181 GCGGACGCCG GAATCAAGGC GCCCTGTGG TCCGTACCC AGGGCGCGGT CTCCGTGCGA
45 6241 CGTCTCGACA CCCCCGCCGA CCCCAGCCGG GCCATGCTCT GGGGCCTCGG CCGCGTCGTC
6301 GCCCTTGAGC ACCCCGAACG CTGGGCGGGC CTCGTGACCT TCCCCGCCCA GCCCGATGCC
6361 GCCGCCCTCG CCCACCTCGT CACCGCACTC TCCGGCGCCA CCGGCGAGGA CCAGATCGCC
6421 ATCCGCACCA CCGGACTCCA CGCCCGCCGC CTCGCCCGCG CACCCTCCA CGGACGTCGG
6481 CCCACCCGCG ACTGGCAGCC CCACGGCACC GTCCTCATCA CCGGCGGCAC CGGAGCCCTC
50 6541 GGCAGCCACG CCGCACGCTG GATGGCCAC CACGGAGCCG AACACCTCCT CCTCGTCAGC
6601 CGCAGCGGCG AACAAGCCCC CGGAGCCACC CAACTACCG CCGAATCAC CGCATCGGGC
6661 GCGCGCGTCA CCATCGCCGC CTGCGACGTC GCCGACCCC ACGCCATGCG CACCCTCCTC
6721 GACGCCATCC CCGCCGAGAC GCCCTCACC GCCGTGCTCC ACACCGCCGG CGCGCTCGAC
6781 GACGGCATCG TGGACACGCT GACCGCCGAG CAGGTCCGGC GGGCCCACCG TGCGAAGGCC
55 6841 GTCGGCGCCT CCGTGCTCGA CGAGCTGACC CCGGACCTCG ACCTCGACGC GTTCGTGCTC
6901 TTCTCGTCCG TGTCGAGCAC TCTGGGCATC CCCGGTCAGG GCAACTACGC CCCGCACAAC
6961 GCCTACCTCG ACGCCCTCGC GGCTCGCCGC CCGGCCACCG GCCGGTCCGC CGTCTCGGTG
7021 GCCTGGGGAC CGTGGGACGG TGGCGGCATG GCCGCCGGTG ACGGCGTGGC CGAGCGGCTG
7081 CGCAACCACG GCGTGCCCGG CATGGACCCG GAACTCGCCC TGGCCGCACT GGAGTCCGCG

7141 CTCGGCCGGG ACGAGACCGC GATCACCGTC GCGGACATCG ACTGGGACCG CTTCTACCTC
7201 GCGTACTCCT CCGGTCGCCC GCAGCCCCTC GTCGAGGAGC TGCCCGAGGT GCGGCGCATC
7261 ATCGACGCAC GGGACAGCGC CACGTCCGGA CAGGGCGGGA GCTCCGCCCA GGGCGCCAAC
7321 CCCCTGGCCG AGCGGCTGGC CGCCGCGGCT CCCGGCGAGC GTACGGAGAT CCTCCTCGGT
5 7381 CTCGTACGGG CGCAGGCCGC CGCCGTGCTC CGGATGCGTT CGCCGGAGGA CGTCGCCGCC
7441 GACCGCGCCT TCAAGGACAT CGGCTTCGAC TCGCTCGCCG GTGTCGAGCT GCGCAACAGG
7501 CTGACCCGGG CGACCGGGCT CCAGCTGCCC GCGACGCTCG TCTTCGACCA CCCGACGCCG
7561 CTGGCCCTCG TGTGCTGCT CCGCAGCGAG TTCCTCGGTG ACGAGGAGAC GCGGACGCC
7621 CCGCGGTCCG CCGCGCTGCC CGCGACTGTC GGTGCCGGTG CCGGCGCCGG CGCCGGCACC
10 7681 GATGCCGACG ACGATCCGAT CGCGATCGTC GCGATGAGCT GCCGCTACCC CGGTGACATC
7741 CGCAGCCCGG AGGACCTGTG GCGGATGCTG TCCGAGGGCG GCGAGGGCAT CACGCCGTTC
7801 CCCACCGACC GCGGCTGGGA CCTCGACGGC CTGTACGACG CCGACCCGGA CGCGCTCGGC
7861 AGGGCGTACG TCCGCGAGGG CGGGTTCCTG CACGACGCGG CCGAGTTCGA CGCGGAGTTC
7921 TTCGGCGTCT CGCCGCGCGA GCGGCTGGCC ATGGACCCGC AGCAGCGGAT GCTCCTGACG
15 7981 ACGTCTTGGG AGGCCTTCGA GCGGGCCGGC ATCGAGCCGG CATCGCTGCG CGGCAGCAGC
8041 ACCGGTGTCT TCATCGGCCT CTCCTACCAG GACTACGCGG CCCGCGTCCC GAACGCCCCG
8101 CGTGGCGTGG AGGGTTACCT GCTGACCGGC AGCAGCCGA GCGTCGCGTC GGGCCGTATC
8161 GCGTACACCT TCGGTCTCGA AGGGCCCGCG ACGACGTCG ACACCGCCTG CTCGTGCTCG
20 8221 CTGACCGCCC TGCACCTGGC GGTGCGGGCG CTGCGCAGCG GCGAGTGCAC GATGGCGCTC
8281 GCCGGTGGCG TGGCGATGAT GCGGACCCCG CACATGTTCG TGGAGTTCAG CCGTCAGCGG
8341 GCGCTCGCCC CGGACGGCCG CAGCAAGGCC TTCTCGGCGG ACGCCGACGG GTTCGGCGCC
8401 GCGGAGGGCG TCGGCCTGCT GCTCGTGGAG CGGCTCTCGG ACGCGCGGCG CAACGGTCAC
8461 CCGGTGCTCG CCGTGGTCCG CGGTACC GCCG GTCAACCAGG ACGGCGCCAG CAACGGGCTG
8521 ACCGCGCCCA ACGGACCCTC GCAGCAGCGG GTGATCCGGC AGGCGCTCGC CGACGCCCGG
25 8581 CTGGCACCCG GCGACATCGA CGCGTCGAG ACGCAGGCA CGGGAACCTC GCTGGGCGAC
8641 CCCATCGAGG CCCAGGGCCT CCAGGCCACG TACGGCAAGG AGCGGCCCGC GGAACGGCCG
8701 CTCGCCATCG GCTCCGTGAA GTCCAACATC GGACACACCC AGGCCGCGGC CGGTGCGGCG
8761 GGCATCATCA AGATGGTCCT CGCGATGCGC CACGGCACCC TGCCGAAGAC CCTCCACGCC
8821 GACGAGCCGA GCCCGCACGT CGACTGGGCG AACAGCGGCC TGGCCCTCGT CACCGAGCCG
30 8881 ATCGACTGGC CGGCCGGCAC CGGTCCGCGC CGCGCCGCGG TCTCCTCCTT CGGCATCAGC
8941 GGGACGAACG CGCACGTCGT GCTGGAGCAG GCGCCGGATG CTGCTGGTGA GGTGCTTGGG
9001 GCCGATGAGG TGCTGAGGT GTCTGAGACG GTAGCGATGG CTGGGACGGC TGGGACCTCC
9061 GAGGTCGCTG AGGGCTCTGA GGCCTCCGAG GCGCCCGCGG CCGCCGGCAG CCGTGAGGCG
9121 TCCCTCCCCG GGCACCTGCC CTGGGTGCTG TCCGCCAAGG ACGAGCAGTC GCTGCGCGGC
35 9181 CAGGCCGCGG CCCTGCACGC GTGGCTGTCC GAGCCCGCCG CCGACCTGTC GGACGCGGAC
9241 GGACCGGCCG GCCTGCGGGA CGTCGGGTAC ACGCTCGCCA CGAGCCGTAC CGCCTTCGCG
9301 CACCGCGCCG CCGTGACCGC CGCCGACCGG GACGGGTTC TGGACGGGCT GGCCACGCTG
9361 GCCCAGGGCG GCACCTCGGC CCACGTCCAC CTGGACACCG CCCGGGACGG CACCACCGCG
9421 TTCTCTTCA CCGGCCAGGG CAGTCAGCGC CCCGGCGCCG GCCGTGAGCT GTACGACCGG
40 9481 CACCCCGTCT TCGCCCGGGC GCTCGACGAG ATCTGCGCCC ACCTCGACGG TCACCTCGAA
9541 CTGCCCCTGC TCGACGTGAT GTTCGCGGCC GAGGGCAGCG CGGAGGCCGC GCTGCTCGAC
9601 GAGACGCGGT ACACGCAAGT CGCGCTGTTT GCCCTGGAGG TCGCGCTCTT CCGGCTCGTC
9661 GAGAGCTGGG GCATGCGGCC GGCCGCACTG CTCGGTCACT CCGTCGGCGA GATCGCCGCC
9721 GCGCACGTCG CCGGTGTGTT CTCGCTCGCC GACGCCGCCG GCCTGGTCGC CGCGCGCGGC
45 9781 CGGCTCATGC AGGAGCTGCC CGCCGGTGGC GCGATGCTCG CCGTCCAGGC CGCGGAGGAC
9841 GAGATCCCGG TGTGGCTGGA GACGGAGGAG CCGTACGCGG GACGTCTGGA CGTCGCCGCC
9901 GTCAACGGCC CCGAGGCCGC CGTCCTGTCC GGCGACGCGG ACGCGGCGCG GGAGGCGGAG
9961 GCGTACTGGT CCGGGCTCGG CCGCAGGACC CGCGCGCTGC GGGTCAGCCA CGCCTTCCAC
50 10021 TCCGCGCACA TGGACGGCAT GCTCGACGGG TTCCGCGCCG TCCTGGAGAC GGTGGAGTTC
10081 CCGCGCCCCCT CCCTGACCGT GGTCTCGAAC GTCACCGGCC TGGCCGCCGG CCCGGACGAC
10141 CTGTGCGACC CCGAGTACTG GGTCCGGCAC GTCCGCGGCA CCGTCCGCTT CCTCGACGGC
10201 GTCCGTGTCC TGCGCGACCT CGGCGTGCGG ACCTGCCTGG AGCTGGGCCG CGACGGGGTC
10261 CTCACCGCCA TGGCGGCCGA CGGCCTCGCG GACACCCCGG CGGATTCCGC TGCCGGCTCC
10321 CCCGTCGGCT CTCCCGCCGG CTCTCCCGCC GACTCCGCCG CCGGCGCGCT CCGGCCCGG
55 10381 CCGCTGCTCG TGGCGCTGCT GCGCCGCAAG CCGTCGGAGA CCGAGACCGT CGCGGACGCC
10441 CTCGGCAGGG CGCACGCCCA CGGCACCGGA CCCGACTGGC ACGCCTGGTT CGCCGGCTCC
10501 GGGGCGCACC GCGTGGACCT GCCACGTAC TCCTTCCGGC GCGACCGCTA CTGGCTGGAC
10561 GCGCCGGCGG CCGACACCGC GGTGGACACC GCGGCGCTCG GTCTCGGCAC CGCCGACCAC
10621 CCGCTGCTCG GCGCCGTGGT CAGCCTTCCG GACCGGGACG GCCTGCTGCT CACCGGCCGC

10681 CTCTCCCTGC GCACCCACCC GTGGCTCGCG GACCACGCCG TCCTGGGGAG CGTCCTGCTC
10741 CCCGGCGCCG CGATGGTCGA ACTCGCCGCG CACGCTGCGG AGTCCGCCGG TCTGCGTGAC
10801 GTGCGGGAGC TGACCCTCCT TGAACCGCTG GTACTGCCCC AGCACGGTGG CGTCGAGCTG
10861 CGCGTGACGG TCGGGGCGCC GGCCGGAGAG CCCGGTGGCG AGTCGGCCGG GGACGGCGCA
5 10921 CGGCCCCTCT CCCTCCACTC GCGGCTCGCC GACGCGCCCC CCGGTACCGC CTGGTCCTGC
10981 CACGCGACCG GTCTGCTGGC CACCGACCGG CCCGAGCTTC CCGTCGCGCC CGACCGTGCG
11041 GCCATGTGGC CGCCGCAGGG CGCCGAGGAG GTGCCGCTCG ACGGTCTCTA CGAGCGGCTC
11101 GACGGGAACG GCCTCGCCTT CGGTCCGCTG TTCCAGGGGC TGAACGCGGT GTGGCGGTAC
11161 GAGGGTGAGG TCTTCGCCGA CATCGCGCTC CCCGCCACCA CGAATGCGAC CGCGCCCGCG
10 11221 ACCGCGAACG GCGGCGGGAG TCGGCGGGCG GCCCCCTACG GCATCCACCC CGCCCTGCTC
11281 GACGCTTCGC TGCACGCCAT CGCGGTGCGC GGTCTCGTCG ACGAGCCCGA GCTCGTCCGC
11341 GTCCCTTCC ACTGGAGCGG TGTACCGTG CACGCGGCCG GTGCCGCGGC GGCCCGGGTC
11401 CGTCTCGCCT CCGCGGGGAC GGACGCCGTC TCGTGTCCC TGACGGACGG CGAGGGACGC
11461 CCGCTGGTCT CCGTGGAACG GCTCACGCTG CGCCCGGTCA CCGCCGATCA GGCGGCGGCG
15 11521 AGCCGCGTCG GGGGGCTGAT GCACCGGGTG GCCTGGCGTC CGTACGCCCT CGCCTCGTCC
11581 GGCGAACAGG ACCCGCACGC CACTTCGTAC GGGCCGACCG CCGTCCTCGG CAAGGACGAG
11641 CTGAAGGTCG CCGCCGCCCT GGAGTCCGCG GCGTTCGAAG TCGGGCTCTA CCGCGACCTG
11701 GCCGCGCTGT CCCAGGACGT GCGGCGCGGC GCGCCGCGCG CCCGTACCGT CCTTGCGCCG
11761 CTGCCCAGCG GTCCCGCCGA CGGCGGCGCG GAGGGTGTAC GGGGCACGGT GGCCCGGACG
20 11821 CTGGAGCTGC TCCAGGCCTG GCTGGCCGAC GAGCACCTCG CGGGCACCCG CCTGCTCCTG
11881 GTCACCCGCG GTGCGGTGCG GGACCCCGAG GGGTCCGGCG CCGACGATGG CGGCGAGGAC
11941 CTGTGCGACG CGGCCGCCTG GGTCTCGTA CCGACCGCGC AGACCGAGAA CCGCGGCCGC
12001 TTCGGCCTTC TCGACCTGGC CGACGACGCC TCGTCTGACC GGACCCTGCC GTCGGTGCTC
12061 TCCGACGCGG GCCTGCGCGA CGAACCGCAG CTCGCCCTGC ACGACGGCAC CATCAGGCTG
25 12121 GCGCGCCTGG CCTCCGTCCG GCGCGAGACC GGCACCGCCG CACCGGCGCT CGCCCGGAG
12181 GGCACGGTCC TGCTGACCGG CCGCACCGGC GGCCTGGGCG GACTGGTTCG CCGGCACGTG
12241 GTGGGCGAGT GGGGCGTACG ACGCCTGCTG CTGGTGAGCC GCGGGGACG GGACGCCCCG
12301 GCGGCCGACG AGCTCGTGCA CGAGCTGGAG GCCCTGGGAG CCGACGTCTC GGTGGCCGCG
12361 TCGGACGTCG CCGACCGCGA AGCCCTCACC GCCGTACTCG ACGCCATCCC CGCCGAACAC
30 12421 CCGCTCACC GGTGCTCCA CACGGCAGGC GTCCTCTCCG ACGGCACCCT CCGTCCATG
12481 ACGACGGAGG ACGTGGAACA CGTACTGCGG CCAAGGTTCG ACGCCGCGTT CCTCCTCGAC
12541 GAACTCACCT CGACGCCCGC ATACGACCTG GCAGCGTTTCG TCATGTTCTC CTCCGCCGCC
12601 GCCGTCTTCG GTGGCGCGGG GCAGGGCGCC TACGCCGCCG CCAACGCCAC CCTCGACGCC
12661 CTCGCCTGGC GCGCCCGGGC AGCCGGACTC CCCGCCCTCT CCCTCGGCTG GGGCCTCTGG
35 12721 GCGGAGACCA GCGGCATGAC CCGCGAGCTC GGCCAGGCGG ACCTGCGCCG GATGAGCCGC
12781 GCGGGCATCG GCGGGATCAG CGACGCCGAG GGCATCGCGC TCCTCGACGC CGCCCTCCGC
12841 GACGACCGCC ACCCGGTCTT GCTGCCCTG CCGCTCGACG CCGCCGGGCT GCGGGACGCG
12901 GCGGGGAACG ACCCGGCCGG AATCCCGGCG CTCTTCCGGG ACGTCGTTCG CGCCAGGACC
12961 GTCCGGGCCC GGCCGTCCGC GGCCTCCGCC TCGACGACAG CCGGGACGGC CGGCACGCCG
40 13021 GGGACGGCGG ACGGCGCGGC GGAACGGCG GCGGTACGC TCGCCGACCG GGCCGCCACC
13081 GTGGACGGGC CCGCACGGCA GCGCCTGCTG CTCGAGTTTCG TCGTCGGCGA GGTGCGCGAA
13141 GTACTCGGCC ACGCCGCGG TCACCGGATC GACGCCGAAC GGGGCTTCCT CGACCTCGGC
13201 TTCGACTCCC TGACCGCCGT CGAACTCCGC AACC GGCTCA ACTCCGCCG TGGCCTCGCC
13261 CTCCCGGCGA CCCTGGTCTT CGACCACCA AGCCCGGCGG CACTCGCCTC CCACCTGGAC
45 13321 GCGGAGCTGC CCGCGGCGC CTCGACAGG GACGGAGCCG GGAACCGGAA CGGGAACGAG
13381 AACGGGACGA CGGCGTCCCG GAGCACCGCC GAGACGGACG CGCTGCTGGC ACAACTGACC
13441 CGCCTGGAAG GCGCCTTGGT GCTGACGGGC CTCTCGGACG CCGCCGGGAG CGAAGAAGTC
13501 CTGGAGCACC TCGGGTCCCT GCGCTCGATG GTCACGGGCG AGACCGGGAC CGGGACCGCG
13561 TCCGGAGCCC CCGACGGCGC CCGGTCCGGC GCGGAGGACC GGCCCTGGG GCGCGGGGAC
50 13621 GGAGCCGGGG GCGGGAGTGA GGACGGCGCG GGAGTGCCGG ACTTCATGAA CGCCTCGGCC
13681 GAGGAACTCT TCGGCCTCCT CGACAGGAC CCCAGCACGG ACTGATCCCT GCGGCACGGT
13741 CGCCTCCCGC CCGGACCCC GTCCCGGGCA CCTCGACTCG AATCACTTCA TGCGCGCCTC
13801 GGGCGCCTCC AGGAACTCAA GGGGACAGCG GTGCCACGGT GAACGAAGAG AAGTACCTCG
13861 ACTACCTGCG TCGTGCCACG GCGGACCTCC ACGAGGCCCG TGGCCGCTC CGCGAGCTGG
55 13921 AGGCGAAGGC GGGCGAGCCG GTGGCGATCG TCGGCATGGC CTGCCGCTG CCGGCGGGCG
13981 TCGCCTCGCC CGAGGACCTG TGGCGGCTGG TGGCCGGCGG CGAGGACGCG ATCTCGGAGT
14041 TCCCCAGGA CCGCGGCTGG GACGTGGAGG GCCTGTACGA CCCGAACCCG GAGGCCACGG
14101 GCAAGAGTTA CGCCCGCGAG GCCGGATTCC TGTACGAGGC GGGCGAGTTC GACGCCGACT
14161 TCTTCGGGAT CTCGCCGCGC GAGGCCCTCG CCATGGACCC GCAGCAGCGT CTCCTCCTGG

14221 AGGCCTCCTG GGAGGCGTTC GAGCACGCCG GGATCCCGGC GGCCACCGCG CGCGGCACCT
14281 CGGTCCGGCGT CTTACCCGGC GTGATGTACC ACGACTACGC CACCCGTCTC ACCGATGTCC
14341 CGGAGGGCAT CGAGGGCTAC CTGGGCACCG GCAACTCCGG CAGTGTGCGC TCGGGCCGCG
14401 TCGCGTACAC GCTTGGCCTG GAGGGGCCGG CCGTCACGGT CGACACCGCC TGCTCGTCCT
5 14461 CGCTGGTCGC CCTGCACCTC GCCGTGCAGG CCCTGCGCAA GGGCGAGGTC GACATGGCGC
14521 TCGCCGGCGG CGTGACGGTC ATGTCGACGC CCAGCACCTT CGTCGAGTTC AGCCGTCAGC
14581 GCGGGCTGGC GCCGGACGGC CGGTGGAAGT CCTTCTCGTC GACGGCCGAC GGCACCAGCT
14641 GGTCCGAGGG CGTCGGCGTC CTCCTCGTCG AGCGCCTGTC CGACGCGCGT CGCAAGGGCC
14701 ATCGGATCCT CGCCGTGGTC CGGGGCACCG CCGTCAACCA GGACGGCGCC AGCAGCGGCC
10 14761 TCACGGCTCC GAACGGGGCCG TCGCAGCAGC GCGTCATCCG ACGTGCCCTG GCGGACGCCC
14821 GGCTCAGGAC CTCCGACGTG GACGTCGTG AGGCCCACGG CACGGGTACG CACTCGGCG
14881 ACCCGATCGA GGCGCAGGCC GTCATCGCCA CGTACGGGCA GGGCCGTGAC GGCGAACAGC
14941 CGCTGCGCCT CGGGTCGTTG AAGTCCAACA TCGGACACAC CCAGGCCGCC GCCGGTGTCT
15001 CCGGCGTGAT CAAGATGGTC CAGGCGATGC GCCACGGCGT CCTGCCGAAG ACGTCCACG
15 15061 TGGAGAAGCC GACGGACCAG GTGGACTGGT CCGCGGGCGC GGTCGAGCTG CTCACCGAGG
15121 CCATGGACTG GCCGGACAAG GCGGACGGCG GACTGCGCAG GGCCGCGGTC TCCTCCTTCG
15181 GCGTCAGCGG GACGAACGCG CACGTCGTGC TCGAAGAGGC CCCGGCGGCC GAGGAGACCC
15241 CTGCCTCCGA GGCGACCCCG GCCGTGAGC CGTCGGTCGG CGCCGGCCTG GTGCCGTGGC
15301 TGGTGTGCGC GAAGACTCCG GCCGCGCTGG ACGCCAGAT CGGACGCCTC GCCGCGTTCG
20 15361 CCTCGCAGGG CCGTACGGAC GCCGCCGATC CCGGCGCGGT CGCTCGCGTA CTGGCCGGCG
15421 GGCGCGCCGA GTTCGAGCAC CGGGCCGTG TGCTCGGCAC CGGACAGGAC GATTTCGCGC
15481 AGGCGCTGAC CGCTCCGGA GACTGATAC GCGGCACGCC CTCGGACGTG GGCCGGGTGG
15541 CGTTCGTGTT CCCCCTCAG GGCACGCACT GGGCCGGGAT GGGCGCCGAA CTCCTCGACG
15601 TGTCGAAGGA GTTCGCGGCG GCCATGGCCG AGTGCAGAG CGCGCTCTCC CGCTATGTGCG
25 15661 ACTGGTCGCT GGAGGCCGTC GTCCGGCAGG CGCCGGGCGC GCCACGCTG GAGCGGGTCC
15721 ACGTCGTCCA GCCCGTGACC TTCGCTGTCA TGGTTTCGCT GGCGAAGGTC TGGCAGCACC
15781 ACGGCGTGAC GCCGAGGCC GTCGTGCGCC ACTCGCAGGG CGAGATCGCC GCCGCGTACG
15841 TCGCCGGTGC CCTCACCTC GACGACGCCG CCCGCGTCGT CACCCTGCGC AGCAAGTCCA
15901 TCGCCGCCCA CCTCGCCGGC AAGGGCGGCA TGATCTCCCT CGCCCTCAGC GAGGAAGCCA
30 15961 CCCGGCAGCG CATCGAGAAC CTCCACGGAC TGTCGATCGC CGCCGTCAAC GGCCCCACCG
16021 CCACCGTGGT TTCGGGCGAC CCCACCCAGA TCCAAGAGCT CGCTCAGGCG TGTGAGGCCG
16081 ACGGGGTCCG CGCACGGATC ATCCCCGTG ACTACGCCTC CCACAGCGCC CACGTCGAGA
16141 CCATCGAGAG CGAACTCGCC GAGGTCCCTG CCGGGCTCAG CCCGCGGACA CCTGAGGTGC
16201 CGTTCTTCTC GACACTCGAA GCGCCTGGA TCACCGAGCC GGTGCTCGAC GGCACCTACT
35 16261 GGTACCGCAA CCTCCGCCAC CGCGTCGGCT TCGCCCCCGC CGTCGAGACC CTCGCCACCG
16321 ACGAAGGCTT CACCCACTTC ATCGAGGTCA GCGCCACCC CGTCCTCACC ATGACCCTCC
16381 CCGAGACCGT CACCGGCCTC GGCACCCTCC GCCGCGAACA GGGAGGCCAG GAGCGTCTGG
16441 TCACCTCACT CGCCGAAGCC TGGACCAACG GCCTCACCAT CGACTGGGCG CCCGTCTCTC
16501 CCACCGCAAC CGGCCACCAC CCCGAGCTCC CCACCTACGC CTTCCAGCGC CGTCACTACT
40 16561 GGCTCCACGA CTCCCCCGC GTCCAGGGCT CCGTGCAGGA CTCCTGGCGC TACCGCATCG
16621 ACTGGAAGCG CCTCGCGGTC GCCGACGCGT CCGAGCGCGC CGGGCTGTCC GGGCGCTGGC
16681 TCGTCGTGCT CCCCAGGAGC CGTTCCGCGG AGGCGGCCCC GGTGCTCGCC GCGCTGTCCG
16741 GCGCCGGCGC CGACCCCGTA CAGCTGGACG TGTCCCGCT GGGCGACCGG CAGCGGCTCG
16801 CCGCGACGCT GGGCGAGGCC CTGGCGGCG CCGGTGGAGC CGTCGACGGC GTCCTCTCGC
45 16861 TGCTCGCGTG GGACGAGAGC GCGCACCCCG GCCACCCCGC CCCCTTCACC CGGGGCACCG
16921 GCGCCACCCT CACCCTGGTG CAGGCGCTGG AGGACGCCGG CGTCGCCGCC CCGCTGTGGT
16981 GCGTGACCCA CGGCGCGGTG TCCGTGCGCC GGGCCGACCA CGTCACCTCC CCCGCCAGG
17041 CCATGGTGTG GGGCATGGGC CGGGTCGCC CCCTGGAGCA CCCCAGCGG TGGGGCGGCC
17101 TGATCGACCT GCCCTCGGAC GCCGACGGG CGGCCCTGGA CCGCATGACC ACGGTCTCTG
50 17161 CCGGCGGTAC GGGTGAGGAC CAGGTGCGCG TACGCGCCTC CGGGCTGCTC GCCCGCCGCC
17221 TCGTCCGCGC CTCCCTCCCG GCGCACGGCA CGGCTTCGCC GTGGTGGCAG GCCGACGGCA
17281 CCGTGCTCGT CACCGGTGCC GAGGAGCCTG CGGCCGCCGA GGCCGACGC CGGCTGGCCC
17341 GCGACGGCGC CGGACACCTC CTCCTCCACA CCACCCCTC CGGCAGCGAA GGCGCCGAAG
17401 GCACCTCCGG TGCCGCCGAG GACTCCGGCC TCGCCGGGCT CGTCGCCGAA CTCGCGGACC
55 17461 TGGGCGCGAC GGCCACCGTC GTGACCTGCG ACCTCACGGA CGCGGAGGCG GCCGCCGGC
17521 TGCTCGCCGG CGTCTCCGAC GCGACCCCG TCAGCGCCGT CCTCCACCTG CCGCCACCG
17581 TCGACTCCGA GCCGCTCGCC GCGACCGAG CCGACGCGCT CGCCCGTGTC GTGACCGCGA
17641 AGGCCACCGC CGCGCTCCAC CTGGACCGCC TCCTGCGGGA GGCCGCGGCT GCCGGAGGCC
17701 GTCCGCCCGT CCTGGTCCTC TTCTCCTCG TCGCCGCGAT CTGGGGCGGC GCCGGTCAGG

17761 GCGCGTACGC CGCCGGTACG GCCTTCCTCG ACGCCCTCGC CGGTCAGCAC CGGGCCGACG
17821 GCCCCACCGT GACCTCGGTG GCCTGGAGCC CCTGGGAGGG CAGCCGCGTC ACCGAGGGTG
17881 CGACCGGGGA GCGGCTGCGC CGCCTCGGCC TGCGCCCCCT CGCCCCCGCG ACGGCGCTCA
17941 CCGCCCTGGA CACCGCGCTC GGCCACGGCG ACACCGCCGT CACGATCGCC GACGTCGACT
5 18001 GGTCGAGCTT CGCCCCCGGC TTCACCACGG CCGGGCCGGG CACCCTCCTC GCCGATCTGC
18061 CCGAGGCGCG CCGCGCGCTC GACGAGCAGC AGTCGACGAC GGCCGCCGAC GACACCGTCC
18121 TGAGCCGCGA GCTCGGTGCG CTCACCGGCG CCGAACAGCA GCGCCGTATG CAGGAGTTGG
18181 TCCGCGAGCA CCTCGCCGTG GTCCTCAACC ACCCTCCCC CGAGGCCGTC GACACGGGGC
18241 GGGCCTTCCG TGACCTCGGA TTCGACTCGC TGACGGCGGT CGAGCTCCGC AACC GCCTCA
10 18301 AGAACGCCAC CGGCCTGGCC CTCCCGGCCA CTCTGGTCTT CGACTACCCG ACCCCCCGGA
18361 CGCTGGCGGA GTTCTCTCTC GCGGAGATCC TGGGCGAGCA GGCCGGTGCC GCGAGCAGC
18421 TTCCGGTGGA CCGCGGGGTC GACGACGAGC CCGTCGCGAT CGTCGGCATG GCGTGCCGCC
18481 TGCCGGGCGG TGTGCGCTCG CCGGAGGACC TGTGGCGGCT GGTGGCCGGC GCGGAGGACG
18541 CGATCTCCGG CTTCCCGCAG GACCGCGGCT GGGACGTGGA GGGGCTGTAC GACCCGACC
15 18601 CCGACGCGTC CGGGCGGACG TACTGCCGTG CCGGTGGCTT CCTCGACGAG GCGGGCGAGT
18661 TCGACGCCGA CTTCTTCGGG ATCTCGCCGC GCGAGGCCCT CGCCATGGAC CCGCAGCAGC
18721 GGCTCCTCCT GGAGACCTCC TGGGAGGCCG TCGAGGACGC CGGGATCGAC CCGACCTCCC
18781 TTCAGGGGCA GCAGGTGCGC GTGTTGCGCG GCACCAACGG CCCCCACTAC GAGCCGCTGC
18841 TCCGCAACAC CGCCGAGGAT CTTGAGGGTT ACGTCGGGAC GGGCAACGCC GCCAGCATCA
20 18901 TGTCGGGCGG TGTCTCGTAC ACCCTCGGCC TGGAGGGCCC GGCCGTCACG GTCGACACCG
18961 CCTGCTCCTC CTCGCTGGTC GCCCTGCACC TCGCCGTGCA GGCCCTGCGC AAGGGCGAAT
19021 GCGGACTGGC GCTCGCGGGC GGTGTGACGG TCATGTGACG GCCACGACG TTCGTGGAGT
19081 TCAGCCGGCA GCGCGGGCTC GCGGAGGACG GCCGGTCGAA GGCGTTCGCC GCGTCGGCGG
19141 ACGGCTTCGG CCCGGCGGAG GCGCTCGGCA TGCTCCTCGT CGAGCGCCTG TCGGACGCC
25 19201 GCCGCAACGG ACACCGTGTG CTGGCGGTGCG TGCGCGGCAG CGCGGTCAAC CAGGACGGCG
19261 CGAGCAACGG CCTGACCGCC CCGAACGGGC CCTCGCAGCA GCGCGTCATC CGGCGCGCGC
19321 TCGCGGACGC CCGACTGACG ACCGCCGACG TGGACGTGCT CGAGGCCAC GGCACGGGCA
19381 CGCGACTCGG CGACCCGATC GAGGCACAGG CCCTCATCGC CACCTACGGC CAGGGGCGCG
19441 ACACCGAACA GCCGCTGCGC CTGGGGTCTG TGAAGTCAA CATCGGACAC ACCAGGCCG
30 19501 CCGCCGGTGT CTCCGGCATC ATCAAGATGG TCCAGGCGAT GCGCCACGGC GTCCTGCCGA
19561 AGACGCTCCA CGTGGACCGG CCGTCGGACC AGATCGACTG GTCGGCGGGC ACGGTGAGC
19621 TGCTCACC GA GGCCATGGAC TGGCCGAGGA AGCAGGAGGG CGGGCTGCGC CGCGCGGCCG
19681 TCTCCTCCTT CGGCATCAGC GGCACGAACG CGCACATCGT GCTCGAAGAA GCCCCGGTCG
19741 ACGAGGACGC CCCGGCGGAC GAGCCGTGCG TCGGCGGTGT GGTGCCGTGG CTCGTGTCCG
35 19801 CGAAGACTCC GGCCGCGCTG GACGCCGAGA TCGGACGCCT CGCCGCGTTC GCCTCGCAGG
19861 GCGTACGGA CGCCGCCGAT CCGGGCGCGG TCGCTCGCGT ACTGGCCGGC GGGCGTGC GC
19921 AGTTCGAGCA CCGGGCGGTC GCGCTCGGCA CCGGACAGGA CGACCTGGCG GCCGCACTGG
19981 CCGCGCCTGA GGGTCTGGTC CGGGGTGTGG CCTCCGGTGT GGGTCGAGTG GCGTTCGTGT
40 20041 TCCCGGGACA GGGCACGCAG TGGGCCGGGA TGGGTGCCGA ACTCCTCGAC GTGTGGAAGG
20101 AGTTCGCGGC GGCCATGGCC GAGTGCGAGG CCGCGCTCGC TCCGTACGTG GACTGGTTCG
20161 TGGAGGCCGT CGTCCGACAG GCCCCGGCG CGCCACGCT GGAGCGGGTC GATGTCGTCC
20221 AGCCCGTGAC GTTCGCGCTC ATGGTCTCGC TGGCGAAGGT CTGGCAGCAC CACGGGGTGA
20281 CCGCGCAAGC CGTCGTGCGC CACTCGCAGG GCGAGATCGC CGCCGCGTAC GTCGCCGGTG
20341 CCCTGAGCCT GGACGACGCC GCTCGTGTGCG TGACCCTGCG CAGCAAGTCC ATCGGCGCCC
45 20401 ACCTCGCGGG CCAGGGCGGC ATGCTGTCCC TCGCGCTGAG CGAGGCGGCC GTTGTGGAGC
20461 GACTGGCCGG GTTCGACGGG CTGTCCGTGCG CCGCCGTCAA CGGGCCTACC GCCACCGTGG
20521 TTTCGGGCGA CCCGACCCAG ATCCAAGAGC TCGCTCAGGC GTGTGAGGCC GACGGGGTCC
20581 GCGCACGGAT CATCCCCGTC GACTACGCCT CCCACAGCGC CCACGTCGAG ACCATCGAGA
20641 GCGAACTCGC CGACGTCCTG GCGGGGTGTG CCCCCAGAC ACCCCAGGTC CCCTTCTTCT
50 20701 CCACCCTCGA AGGCGCCTGG ATCACC GAAC CCGCCCTCGA CGGCGGCTAC TGGTACCGCA
20761 ACCTCCGCCA TCGTGTGGGC TTCGCCCCGG CCGTCGAAC CCTGGCCACC GACGAAGGCT
20821 TCACCCACTT CGTCGAGGTC AGCGCCACC CCGTCCTCAC CATGGCCCTG CCCGAGACCG
20881 TCACCGGCCT CGGCACCTC CGCCGTGACA ACGGCGGACA GCACCGCCTC ACCACCTCCC
20941 TCGCCGAGGC CTGGGCCAAC GGCTCACCG TCGACTGGGC CTCTCTCCTC CCCACCACGA
55 21001 CCACCCACCC CGATCTGCC ACCTACGCT TCCAGACCGA GCGCTACTGG CCGCAGCCCG
21061 ACCTCTCCGC CGCCGGTGAC ATCACCCTCG CCGGTCTCGG GGCGGCCGAG CACCCGCTGC
21121 TCGGCGCGGC CGTGGCGCTC GCGGACTCCG ACGGCTGCCT GCTCACGGGG AGCCTCTCCC
21181 TCCGTACGCA CCCCTGGCTG GCGGACCACG CCGTGGCCGG CACCGTGCTG CTGCCGGGAA
21241 CGGCGTTCGT GGAGCTGGCG TTCCGAGCCG GGGACCAGGT CCGTTGCGAT CTGGTTCGAGG

21301 AGCTCACCCT CGACGCGCCG CTCGTGCTGC CCCGTCTGTG CGCGGTCCGT GTGCAGCTGT
21361 CCGTCGGCGC GAGCGACGAG TCCGGGCGTC GTACCTTCGG GCTCTACGCG CACCCGGAGG
21421 ACGCGCCGGG CGAGGCGGAG TGGACGCGGC ACGCCACCGG TGTGCTGGCC GCCCGTGGCG
21481 ACCGCACCGC CCCCGTCGCC GACCCGGAGG CCTGGCCGCC GCCGGGCGCC GAGCCGGTGG
5 21541 ACGTGGACGG TCTGTACGAG CGCTTCGCGG CGAACGGCTA CGGCTACGGC CCCCTCTTCC
21601 AGGGCGTCCG TGGTGTCTGG CGGCGTGGCG ACGAGGTGTT CGCCGACGTG GCCCTGCCGG
21661 CCGAGGTCGC CGGTGCCGAG GGC GCGCGGT TCGGCCTTCA CCCGGCGCTG CTCGACGCCG
21721 CCGTGCAGGC GGCCGGTGGC GCGGGGCGGT TCGGCGCGGG CACGCGGCTG CCGTTCGCCT
21781 GGAGCGGGAT CTCCCTGTAC GCGGTGCGCG CCACCGCCCT CCGCGTGGCG CTGGCCCCCG
10 21841 CCGGCCCGGA CACGGTGTCC GTGAGCGCCG CCGACTCCTC CGGGCAGCCG GTGTTCGCCG
21901 CGGACTCCCT CACGGTGTCT CCCGTCGACC CCGCGCAGCT GGCGGCCTTC AGCGACCCGA
21961 CTCTGGACGC GCTGCACCTG CTGGAGTGGG CCGCCTGGGA CGGTGCCGCG CAGGCCCTGC
22021 CCGGCGCGGT CGTGCTGGGC GCGGACGCCG ACGGTCTCGC CGCGGCGCTG CGCGCCGGTG
22081 GCACCGAGGT CCTGTCTTTC CCGGACCTTA CGGACCTGGT GGAGGCCGTC GACCGGGGCG
15 22141 AGACCCCGGC CCCGGCGACC GTCTTGGTGG CCTGCCCGCG CGCCGGCCCC GGTGGGCCGG
22201 AGCATGTCCG CGAGGCCCTG CACGGGTGCG TCGCGTGTAT GCAGGCCTGG CTGGCCGACG
22261 AGCGGTTCAC CGATGGGCGC CTGGTGTCTG TGACCCGCGA CGCGGTGCGC GCCCGTCCG
22321 GCGACGGCCT GCGGTCCACG GGACAGGCCG CCGTCTGGGG CCTCGGCCGG TCCGCGCAGA
22381 CCGAGAGCCC GGGCCGGTTC GTCTTGTCTG ACCTCGCCGG GGAAGCCCGG ACGGCCGGGG
20 22441 ACGCCACCGC CGGGGACGGC CTGACGACCG GGGACGCCAC CGTCGGCGGC ACCTCTGGAG
22501 ACGCCGCCCT CGGCAGCGCC CTCGCGACCG CCCTCGGCTC GGGCGAGCCG CAGCTCGCCC
22561 TCCGGGACGG GCGGCTCCTC GTACCCCGCC TGGCGCGGGC CGCCGCGCCC GCCGCGGCCG
22621 ACGGCCCTCGC CGCGGCCGAC GGCTCGCCG CTCTGCCGCT GCCCGCCGCT CCGGCCCTCT
22681 GCGGTCTGGA GCCCGGTACG GACGGCAGCC TGGAGAGCCT CACGGCGGGC CCCGGCGACG
25 22741 CCGAGACCCT CGCCCCGGAG CCGCTCGGCC CGGGACAGGT CCGCATCGCG ATCCGGGCCA
22801 CCGGTCTCAA CTTCGCGAC GTCTTGTATG CCCTCGGCAT GTACCCCGAT CCGGCGCTGA
22861 TGGGCACCGA GGGAGCCGGC GTGGTCACCG CGACCGGCC CGGCGTCACG CACCTCGCCC
22921 CCGGCGACCG GGTCATGGGC CTGCTCTCCG GCGCGTACGC CCCGGTCTGC GTGGCGGACG
22981 CGCGGACCGT CGCGCGGATG CCCGAGGGGT GGACGTTCTG CCAGGGCGCC TCCGTGCCGG
30 23041 TGGTGTTCCT GACGGCCGTC TACGCCCTGC GCGACCTGGC GGACGTCAAG CCCGGCGAGC
23101 GCCTCCTGGT CCACTCCGCC GCGGTGGCG TGGGCATGGC CGCCGTGCAG CTCGCCCGGC
23161 ACTGGGGCGT GGAGGTCCAC GGCACGGCGA GTCACGGGAA GTGGGACGCC CTGCGCGCGC
23221 TCGGCCTGGA CGACGCGCAC ATCGCCTCCT CCCGACCCCT GGACTTCGAG TCCGCGTTCC
23281 GTGCCGCTTC CGGCGGGGCG GGCATGGACG TCGTACTGAA CTCGCTCGCC CGCGAGTTCT
35 23341 TCGACGCCTC GCTGCGCCTG CTCGGGCGCG GCGGCCGTT CGTGGAGATG GGAAGACCG
23401 ACGTCCGCGA CGCGGAGCGG GTCGCCGCCG ACCACCCCGG TGTCGGCTAC CGCGCCTTCG
23461 ACCTGGGCGA GGCCGGGCCG GAGCGGATCG GCGAGATGCT CGCCGAGGTC ATCGCCCTCT
23521 TCGAGGACGG GTGCTCCGG CACCTGCCCG TCACGACCTG GGACGTGCGC CGGGCCCGCG
23581 ACGCCTTCCG GCACGTACAG CAGGCCCGCC ACACGGGCAA GGTCGTCTTC ACGATGCCGT
40 23641 CCGGCCCTCGA CCCGGAGGGT ACGGTCCTGC TGACCGGCGG CACCGGTGCG CTGGGGGGCA
23701 TCGTGGCCCG GCACGTGGTG GGCGAGTGGG GCGTACGACG CCTGCTGCTC GTGAGCCGGC
23761 GGGGCACGGA CGCCCCGGGC GCCGGCGAGC TCGTGCACGA GCTGGAGGCC CTGGGAGCCG
23821 ACGTCTCGGT GGCCGCGTGC GACGTGCGCG ACCGCGAAGC CCTACCGCC GTACTCGACT
23881 CGATCCCCGC CGAACACCCG CTCACCGCGG TCGTCCACAC GGCAGGCGTC CTCTCCGACG
45 23941 GCACCCCTCCC CTCGATGACA GCGGAGGATG TGGAACACGT ACTGCGTCCC AAGGTGACG
24001 CCGCGTTCCT CCTCGACGAA CTCACCTCGA CGCCCGGCTA CGACCTGGCA GCGTTCGTCA
24061 TGTTCTCCTC CGCCGCCGCC GTCTTCGGTG GCGCGGGGCA GGGCGCCTAC GCCGCCGCCA
24121 ACGCCACCCT CGACGCCCTC GCCTGGCGCC GCCGGACAGC CGGACTCCCC GCCCTCTCCC
24181 TCGGCTGGGG CCTCTGGGCC GAGACCAGCG GCATGACCGG CGGACTCAGC GACACCGACC
50 24241 GCTCGCGGCT GGCCCGTTCC GGGGCGACGC CCATGGACAG CGAGCTGACC CTGTCCCTCC
24301 TGGACGCGGC CATGCGCCGC GACGACCCGG CGCTCGTCCC GATCGCCCTG GACGTGCGCG
24361 CGCTCCGCGC CCAGCAGCGC GACGGCATGC TGGCGCCGCT GCTCAGCGGG CTCACCCGCG
24421 GATCGCGGGT CGGCGGCGCG CCGGTCAACC AGCGCAGGGC AGCCGCCGGA GGCGCGGGCG
24481 AGGCGGACAC GGACCTCGGC GGGCGGCTCG CCGCGATGAC ACCGGACGAC CGGGTCGCGC
55 24541 ACCTGCGGGA CCTCGTCCGT ACGCACGTGG CGACCGTCTT GGGACACGGC ACCCCGAGCC
24601 GGGTGGACCT GGAGCGGGCC TTCCGCGACA CCGGTTTCGA CTCGCTCACC GCCGTGGAAC
24661 TCCGCAACCG TCTCAACGCC GCGACCGGGC TGCGGCTGCC GGCCACGCTG GTCTTCGACC
24721 ACCCCACCCC GGGGGAGCTC GCCGGGCACC TGCTCGACGA ACTCGCCACG GCCGCGGGCG
24781 GGTCTGGGC GGAAGGCACC GGGTCCGGAG ACACGGCCTC GGCGACCGAT CGGCAGACCA

24841 CGGCGGCCCT CGCCGAATC GACCGGCTGG AAGGCGTGCT CGCCTCCCTC GCGCCCGCCG
24901 CCGGCGGCCG TCCGGAGCTC GCCGCCCGGC TCAGGGCGCT GGCCGCGGCC CTGGGGGACG
24961 ACGGCGACGA CGCCACCGAC CTGGACGAGG CGTCCGACGA CGACCTCTTC TCCTTCATCG
25021 ACAAGGAGCT GGGCGACTCC GACTTCTGAC CTGCCCCGACA CCACCGGCAC CACCGGCACC
5 25081 ACCAGCCCCC CTCACACACG GAACACGGAA CGGACAGGCG AGAACGGGAG CCATGGCGAA
25141 CAACGAAGAC AAGCTCCGCG ACTACCTCAA GCGCGTCACC GCCGAGCTGC AGCAGAACAC
25201 CAGGCGTCTG CGCGAGATCG AGGGACGCAC GCACGAGCCG GTGGCGATCG TGGGCATGGC
25261 CTGCCGCCTG CCGGGCGGTG TCGCCTCGCC CGAGGACCTG TGGCAGCTGG TGGCCGGGGA
25321 CCGGGACGCG ATCTCGGAGT TCCCGCAGGA CCGCGGCTGG GACGTGGAGG GGCTGTACGA
10 25381 CCGCGACCCG GACGCGTCCG GCAGGACGTA CTGCCGGTCC GCGCGATTCC TGCACGACGC
25441 CCGCGAGTTC GACGCCGACT TCTTCGGGAT CTCGCCGCGC GAGGCCCTCG CCATGGACCC
25501 GCAGCAGCGA CTGTCCCTCA CCACCGCGTG GGAGGCGATC GAGAGCGCGG GCATCGACCC
25561 GACGGCCCTG AAGGGCAGCG GCCTCGGCGT CTTCTGTCGGC GGCTGGCACA CCGGCTACAC
25621 CTCGGGGCAG ACCACCGCCG TGCAGTCGCC CGAGCTGGAG GGCCACCTGG TCAGCGGCGC
15 25681 GGCGCTGGGC TTCCTGTCCG GCCGTATCGC GTACGTCTTC GGTACGGACG GACCGGCCCT
25741 GACCGTGGAC ACGGCCTGCT CGTCTCGCT GGTCCGCTCG CACCTCGCCG TGCAGGCCCT
25801 CCGCAAGGGC GAGTGCAGCA TGGCCCTCGC CCGTGGTGTG ACGGTCATGC CCAACGCGGA
25861 CCTGTTCTGT CAGTTCAGCC GGCAGCGCGG GCTGGCCGCG GACGGCCGGT CGAAGGCGTT
25921 CGCCACCTCG GCGGACGGCT TCGGCCCCGC GGAGGGCGCC GGAGTCCTGC TGGTGGAGCG
20 25981 CCTGTCCGAC GCCCGCCGCA ACGGACACCG GATCCTCGCG GTCGTCCGCG GCAGCGCGGT
26041 CAACCAGGAC GGCGCCAGCA ACGGCCTCAC GGCTCCGCAC GGGCCCTCCC AGCAGCGCGT
26101 CATCCGACGG GCCCTGGCGG ACGCCCGGCT CGCGCCGGGT GACGTGGACG TCGTCGAGGC
26161 GCACGGCACG GGCACGCGGC TCGGCGACCC GATCGAGGCG CAGGCCCTCA TCGCCACCTA
26221 CCGCCAGGAG AAGAGCAGCG AACAGCCGCT GAGGCTGGGC GCGTTGAAGT CGAACATCGG
25 26281 GCACACGCAG GCCGCGGGCCG GTGTGCGAGG TGTCATCAAG ATGGTCCAGG CGATGCGCCA
26341 CCGACTGCTG CCGAAGACGC TGCACGTCGA CGAGCCCTCG GACCAGATCG ACTGGTCGGC
26401 GGGCACGGTG GAACTCCTCA CCGAGGCCGT CGACTGGCCG GAGAAGCAGG ACGGCGGGCT
26461 GCGCCGCGCG GCTGTCTCCT CCTTCGGCAT CAGCGGGACG AACGCGCACG TCGTCTTGGA
26521 GGAGGCCCCG GCGGTCGAGG ACTCCCCGGC CGTCGAGCCG CCGGCCGGTG GCGGTGTGGT
30 26581 GCCGTGGCCG GTGTCCGCGA AGACTCCGGC CGCGCTGGAC GCCCAGATCG GGCAGCTCGC
26641 CGCGTACGCG GACGGTCGTA CGGACGTGGA TCCGGCGGTG GCCGCCCGCG CCCTGGTCGA
26701 CAGCCGTACG GCGATGGAGC ACCGCGCGGT CGCGGTGCGC GACAGCCGGG AGGCACTGCG
26761 GGACGCCCTG CCGATGCCGG AAGGACTGGT ACGCGGCACG TCCTCGGACG TGGGCCGGGT
26821 GGCGTTCGTC TTCCCCGGCC AGGGCACGCA GTGGGCGCGC ATGGGCGCCG AACTCCTTGA
35 26881 CAGCTACCG GAGTTCGCTG CCTCGATGGC CGAATGCGAG ACCGCGCTCT CCGCTACGT
26941 CGACTGGTCT CTTGAAGCCG TCGTCCGACA GGAACCCGGC GCACCCACGC TCGACCGCGT
27001 CGACGTGCTC CAGCCCGTGA CCTTCGCTGT CATGGTCTCG CTGGCGAAGG TCTGGCAGCA
27061 CCACGGCATC ACCCCCCAGG CCGTCGTCGG CCACTCGCAG GCGGAGATCG CCGCCGCGTA
27121 CGTCGCCGGT GCACTACCC TCGACGACGC CGCCCGCGTC GTCACCCTGC GCAGCAAGTC
40 27181 CATCGCCGCC CACCTCGCCG GCAAGGGCGG CATGATCTCC CTCGCCCTCG ACGAGGCGGC
27241 CGTCTGAAG CCACTGAGCG ACTTCGACGG ACTCTCCGTC GCCGCCGTCA ACGGCCCCAC
27301 CGCCACCGTC GTCTCCGGCG ACCCGACCCA GATCGAGGAA CTCGCCCGCA CCTGCGAGGC
27361 CGACGGCGTC CGTGCGCGGA TCATCCCGGT CGACTACGCC TCCCACAGCC GGCAGGTCGA
27421 GATCATCGAG AAGGAGCTGG CCGAGGTCTT CGCCGGAATC GCGCCGAGG CTCCGCACGT
45 27481 GCCGTTCTTC TCCACCTCG AAGGCACCTG GATCACCGAG CCGGTGCTCG ACGGCACCTA
27541 CTGGTACCGC AACCTGCGCC ATCGCGTGGG CTTCCGCCCC GCGGTGGAGA CCTTGGCGGT
27601 TGACGGCTTC ACCCACTTCA TCGAGGTCAG CGCCACCCC GTCCTACCA TGACCCTCCC
27661 CGAGACCGTC ACCGGCCTCG GCACCCTCCG CCGCGAACAG GGAGGCCAGG AGCGTCTGGT
27721 CACCTCACTC GCCGAAGCCT GGGCCAACGG CCTCACCATC GACTGGGCGC CCATCCTCCC
50 27781 CACCGCAACC GGCCACCACC CCGAGCTCCC CACCTACGCC TTCCAGACCG AGCGCTTCTG
27841 GCTGCAGAGC TCCGCGCCCA CCAGCGCCG CGACGACTGG CGTTACCGCG TCGAGTGGAA
27901 GCCGCTGACG GCCTCCGGCC AGGCGGACCT GTCCGGGCGG TGGATCGTCG CCGTCGGGAG
27961 CGAGCCAGAA GCCGAGCTGC TGGGCGCGCT GAAGGCCGCG GGAGCGGAGG TCGACGTACT
28021 GGAAGCCGGG GCGGACGACG ACCGTGAGGC CCTCGCCGCC CGGCTACCG CACTGACGAC
55 28081 CCGCGACGGC TTCACGGCG TGGTCTCGCT CCTCGACGAC CTCGTGCCAC AGGTGCGCTG
28141 GGTGCAGGCA CTCGGCGACG CCGGAATCAA GCGGCCCTG TGGTCCGTCA CCCAGGGCGC
28201 GGTCTCCGTC GGACGTCTCG ACACCCCCGC CGACCCCGAC CGGGCCATGC TCTGGGGCCT
28261 CCGCCGCGTC GTCGCCCTTG AGCACCCCGA ACGCTGGGCC GGCCTCGTCG ACCTCCCCGC
28321 CCAGCCCGAT GCCGCCGCC TCGCCACCT CGTCACCGCA CTCTCCGGCG CCACCGCGCA

28381 GGACCAGATC GCCATCCGCA CCACCGGACT CCACGCCCGC CGCCTCGCCC GCGCACCCCT
28441 CCACGGACGT CGGCCACCC GCGACTGGCA GCCCCACGGC ACCGTCCTCA TCACCGGCGG
28501 CACCGGAGCC CTCGGCAGCC ACGCCGCACG CTGGATGGCC CACCACGGAG CCGAACACCT
28561 CCTCCTCGTC AGCCGCAGCG GCGAACAAGC CCCCAGGAGC ACCCAACTCA CCGCCGAACCT
5 28621 CACCGCATCG GCGGCCCGCG TCACCATCGC CGCCTGCGAC GTCGCCGACC CCCACGCCAT
28681 GCGCACCCCTC CTCGACGCCA TCCCCGCCGA GACGCCCTC ACCGCCGTCG TCCACACCGC
28741 CCGCGCACCG GCGGCGGATC CGCTGGACGT CACCGGCCCG GAGGACATCG CCCGCATCCT
28801 GGGCGCGAAG ACGAGCGGCG CCGAGGTCCT CGACGACCTG CTCCGCGGCA CTCCGCTGGA
28861 CGCCTTCGTC CTCTACTCCT CGAACGCCCG GGTCTGGGGC AGCGGCAGCC AGGGCGTCTA
10 28921 CGCGGCGGCC AACGCCACC TCGACGCGCT CGCCGCCCGG CGCCGCGCCC GGGGCGAGAC
28981 GCGGACCTCG GTCGCTGGG GCCTCTGGGC CGGCGACGGC ATGGGCCGGG GCGCCGACGA
29041 CGCGTACTGG CAGCGTCGCG GCATCCGTCC GATGAGCCCC GACCGCGCCC TGGACGAACCT
29101 GGCCAAGGCC CTGAGCCACG ACGAGACCTT CGTCGCCGTG GCCGATGTCG ACTGGGAGCG
29161 GTTCGCGCCC GCGTTCACGG TGTCCCGTCC CAGCCTTCTG CTCGACGGCG TCCCGGAGGC
15 29221 CCGGCAGGCG CTCGCCGCAC CCGTCGGTGC CCGGGCTCCC GCGACGCGC CCGTGGCGCC
29281 GACCGGGCAG TCGTCGGCGC TGGCCGCGAT CACCGCGCTC CCCGAGCCCG AGCGCCGGCC
29341 GCGGCTCCTC ACCCTCGTCC GTACCCACGC GCGGCGCGTA CTCGGCCATT CCTCCCCGA
29401 CCGGGTGGCC CCGGCGCGTG CCTTCACCGA GCTCGGCTTC GACTCGCTGA CCGCCGTGCA
29461 GCTCCGCAAC CAGCTCTCCA CCGTGGTCCG CAACAGGCTC CCCGCCACCA CCGTCTTCGA
20 29521 CCACCCGACG CCCGCCGCAC TCGCCGCGCA CCTCCACGAG GCGTACCTCG CACCGGCCGA
29581 GCCGGCCCCG ACGGACTGGG AGGGGCGGGT GCGCCGGGCC CTGGCCGAAC TGCCCTTCGA
29641 CCGGCTGCGG GACGCGGGGG TCCTCGACAC CGTCCTGCGC CTCACCGGCA TCGAGCCCGA
29701 GCCGGGTTCG GCGGTTCCG ACGGCGGCGC CGCCGACCCT GGTGCGGAGC CGGAGGCGTC
29761 GATCGACGAC CTGGACGCGG AGGCCCTGAT CCGGATGGCT CTCGGCCCCC GTAACACCTG
25 29821 ACCCGACCGC GGTCTGCCC CACGCGCCGC ACCCGCGCA TCCCGCGCAC CACCGCCCCC
29881 CACACGCCCA CAACCCATC CACGAGCGGA AGACCACACC CAGATGACGA GTTCCAACGA
29941 ACAGTTGGTG GACGCTCTGC GCGCCTCTCT CAAGGAGAAC GAAGAACTCC GGAAAGAGAG
30001 CCGTCGCCGG GCCGACCGTC GGCAGGAGCC CATGGCGATC GTCGGCATGA GCTGCCGGTT
30061 CGCGGGCGGA ATCCGGTCCC CCGAGGACCT CTGGGACGCC GTCGCCGCGG GCAAGGACCT
30 30121 GGTCTCCGAG GTACCGGAGG AGCGCGGCTG GGACATCGAC TCCCTCTACG ACCCGGTGCC
30181 CCGGCGCAAG GGCACGACGT ACGTCCGCAA CGCCGCGTTC CTCGACGACG CCGCCGGATT
30241 CGACGCGGCC TTCTTCGGGA TCTCGCCGCG CGAGGCCCTC GCCATGGACC CGCAGCAGCG
30301 GCAGCTCCTC GAAGCCTCCT GGGAGGTCTT CGAGCGGGCC GGCATCGACC CCGCGTCGGT
30361 CCGCGGCACC GACGTCGGCG TGTACGTGGG CTGTGGCTAC CAGGACTACG CGCCGGACAT
35 30421 CCGGGTCGCC CCCGAAGGCA CCGGCGGTTA CGTCGTCACC GGCAACTCCT CCGCCGTGGC
30481 CTCCGGGCGC ATCGCGTACT CCCTCGGCCCT GGAGGGACCC GCCGTGACCG TGGACACGGC
30541 GTGCTCCTCT TCGCTCGTCG CCCTGCACCT CGCCCTGAAG GGCCTGCGGA ACGGCGACTG
30601 CTCGACGGA CTCGTGGGCG GCGTGGCCGT CCTCGCGACG CCGGGCGCGT TCATCGAGTT
30661 CAGCAGCCAG CAGGCCATGG CCGCCGACGG CCGGACCAAG GGCTTCGCTC CCGCGGCGGA
40 30721 CGGCCTCGCC TGGGGCGAGG GCGTCGCCGT ACTCCTCCTC GAACGGCTCT CCGACGCGCG
30781 GCGCAAGGGC CACCGGGTCC TGGCCGTCTG GCGCGGCAGC GCCATCAACC AGGACGGCGC
30841 GAGCAACGGC CTCACGGCTC CGCACGGGCC CTCCAGCAG CGCCTGATCC GCCAGGCCCT
30901 GGCCGACGCG CGGCTCACGT CGAGCGACGT GGACGTCGTG GAGGGCCACG GCACGGGGAC
30961 CCGTCTCGGC GACCCGATCG AGGCGCAGGC GCTGCTCGCC ACGTACGGGC AGGGGCGCGC
45 31021 CCCGGGGCAG CCGCTGCGGC TGGGGACGCT GAAGTCGAAC ATCGGGCACA CGCAGGCCGC
31081 TTCGGGTGTC GCCGGTGTCA TCAAGATGGT GCAGGCGCTG CGCCACGGGG TGCTGCCGAA
31141 GACCCTGCAC GTGGACGAGC CGACGGACCA GGTCGACTGG TCGGCCGGTT CCGTTCGAGCT
31201 GCTCACCAGG GCCGTGGACT GGCCGGAGCG GCCGGGCCGG CTCCGCCGGG CGGGCGTCTC
31261 CGCGTTCGGC GTGGGCGGGA CGAACGCGCA CGTCGTCCTG GAGGAGGCC CGGCGGTCTC
50 31321 GGAGTCCCCT GCCGTGAGC CGCCGGCCGG TGGCGGCGTG GTGCCGTGGC CCGTGTCCGC
31381 GAAGACCTCG GCCGCACTGG ACGCCAGAT CCGGCAGCTC GCCGCATACG CGGAAGACCG
31441 CACGGACGTG GATCCGGCGG TGGCCGCCCG CGCCCTGGTC GACAGCCGTA CCGCGATGGA
31501 GCACCGCGCG GTCGCGGTG GCGACAGCCG GGAGGCACTG CCGGACGCCC TGCGGATGCC
31561 GGAAGGACTG GTACGGGGCA CCGTCACCGA TCCGGGCCGG GTGGCGTTCT TCTTCCCCGG
55 31621 CCAGGGCAG CAGTGGGCCG GCATGGGCGC CGAACTCCTC GACAGCTCAC CCGAATTCTC
31681 CGCCGCCATG GCCGAATGCG AGACCGCACT CTCCCCGTAC GTCGACTGGT CTCTCGAAGC
31741 CGTCGTCCGA CAGGCTCCCA GCGCACCGAC ACTCGACCGC GTCGACGTCG TCCAGCCCGT
31801 CACCTTCGCC GTCATGGTCT CCCTCGCCAA GGTCTGGCAG CACCACGGCA TCACCCCGGA
31861 GCGCGTCATC GGCCACTCCC AGGGCGAGAT CGCCGCCGCG TACGTCGCCG GTGCCCTCAC

31921 CCTCGACGAC GCCGCTCGTG TCGTGACCCT CCGCAGCAAG TCCATCGCCG CCCACCTCGC
31981 CGGCAAGGGC GGCATGATCT CCCTCGCCCT CAGCGAGGAA GCCACCCGGC AGCGCATCGA
32041 GAACCTCCAC GGAATGTCGA TCGCCGCCGT CAACGGGCCT ACCGCCACCG TGGTTTCGGG
5 32101 CGACCCACC CAGATCCAAG AACTTGCTCA GGCGTGTGAG GCCGACGGCA TCCGCGCAGC
32161 GATCATCCCC GTCGACTACG CCTCCACAG CGCCACGTC GAGACCATCG AGAACGAACT
32221 CGCCGACGTC CTGGCGGGGT TGTCCTCCCA GACACCCAG GTCCCTTCT TCTCCACCCT
32281 CGAAGGCACC TGGATCACC GACCCGCCCT CGACGGCGGC TACTGGTACC GCAACCTCCG
32341 CCATCGTGTG GGCTTCGCC CGGCCGTCGA GACCTCGCC ACCGACGAAG GCTTCACCCA
10 32401 CTTATCGAG GTCAGCGCC ACCCGTCCT CACCATGACC CTCCCGACA AGGTCACCGG
32461 CCTGGCCACC CTCCGACCG AGGACGGCG ACAGCACCG CTCACCACCT CCCTTGCCGA
32521 GGCTGGGCC AACGGCCTCG CCCTCGACT GGCTCCCTC CTGCCGCCA CGGGCGCCCT
32581 CAGCCCGCC GTCCCGACC TCCCGACGTA CGCTTCCAG CACCGCTCGT ACTGGATCAG
32641 CCCCAGGGT CCGCGGAGG CGCCCGCGCA CACCGCTTCC GGGCGCGAGG CCGTCGCCGA
15 32701 GACGGGGCTC GCGTGGGGCC CGGGTGCCGA GGACCTCGAC GAGGAGGGCC GGCGCAGCGC
32761 CGTACTCGCG ATGGTGATGC GGCAGGCGGC CTCCGTGCTC CGGTGCGACT CGCCGAAGA
32821 GGTCCCGTC GACCGCCCGC TCGGGGAGAT CGGCTTCGAC TCGTGACCG CCGTCGACTT
32881 CCGCAACCGC GTCAACGGC TGACCGGTCT CCAGCTGCCG CCCACCGTCG TGTTCGAGCA
32941 CCGACGCCG GTCGCGCTCG CCGAGCGCAT CAGCGACGAG CTGGCCGAGC GGAACGGGC
20 33001 CGTCGCCGAG CCGTCGGATC ACGAGCAGG GGAGGAGGAG AAGGCCGCCG CTCCGGCGGG
33061 GGCCCGCTCC GGGCGCGACA CCGCGCCGG CGCCGGGATG TTCCCGCCC TGTTCGGCA
33121 GGCCGTGGAG GACGACCGT ACGCGAGTT CCTCGACGTC CTCGCCAAG CCTCCGCTT
33181 CCGCCCGCAG TTCGCTCGC CCGAGGCCTG CTCGGAGCGG CTCGACCCGG TGCTGCTCGC
33241 CGGCGGTCCG ACGGACCGGG CGGAAGGCCG TGCCGTCTC GTCGGCTGCA CCGCACCGC
25 33301 GGCGAACGGC GGCCCGCACG AGTTCCTGCG GCTCAGACC TCCTTCCAGG AGGAGCGGGA
33361 CTTCTCGCC GTACCTCTCC CCGGCTACGG CACGGGTACG GGCACCGCA CGGCCCTCT
33421 CCGGCGGAT CTCGACACCG CGCTCGACG CCAGGCCCGG GCGATCCTCC GGGCGCGCG
33481 GGACGCCCG GTCGTCTGTC TCGGGCACTC CGGCGCGCC CTGCTCGCGC ACGAGCTGGC
33541 CTTCCGCTG GAGCGGGCGC ACGCGCGCC GCCGGCCGG ATCGTCTGG TCGACCCCTA
30 33601 TCCGCCGGC CATCAGGAG CCATCGAGG GTGGAGCAG CAGCTGGCG AGGGCCTGTT
33661 CGCGGGCGAG CTGGAGCCGA TGTCCGATG CGGCTGCTG GCCATGGGCC GTACGCGCG
33721 GTTCTCGCC GGCCCGCGG CGGGCCGCG CAGCGCGCC GTGCTTCTG TCCGTGCTC
33781 CGAACCGCTG GCGGACTGGC AGGAGGAGCG GGGCGACTGG CGTGCCCACT GGGACCTTCC
33841 GCACACCGTC GCGGACGTG CCGGCGACCA CTTACGATG ATCGGGGACC ACGCGCCGGC
33901 CGTCGCCGAG GCCGTCTCT CCTGGCTCGA CGCCATCGAG GGCATCGAG GGGCGGGCAA
35 33961 GTGACCGACA GACCTCTGAA CGTGGACAG GGAATGTGGA TCCGGCGCTT CCACCCCGCG
34021 CCGAACAGCG CGGTGCGGCT GGTCTGCCT CCGCACGCCG GCGGCTCCG CAGCTACTTC
34081 TTCCGCTTCT CGGAGGAGCT GCACCCCTCC GTCGAGGCC TGTCGGTGCA GTATCCGGGC
34141 CGCCAGGACC GCGGTGCCGA GCCGTGTCTG GAGAGCGTC AGGAGCTCGC CGAGCATGTG
40 34201 GTCGCGGCA CCGAACCTG GTGGCAGGAG GGCCGGCTGG CCTTCTTCGG GCACAGCCTC
34261 GGCGCCTCCG TCGCCTTCGA GACGGCCCGC ATCCTGGAAC AGCGGCACGG GGTACGGCCC
34321 GAGGGCCTGT ACGTCTCCGG TCGGCGCGCC CCGTCGCTGG CGCCGGACCG GCTCGTCCAC
34381 CAGCTGGACG ACCGGGCGTT CCTGGCCGAG ATCCGGCGGC TCAGCGGCAC CGACGAGCGG
34441 TTCTCCAGG ACGACGAGCT GCTGCGGCTG GTGCTGCCCC CGCTGCGCAG CGACTACAAG
45 34501 GCGGCGGAGA CGTACCTGCA CCGGCCGTCC GCCAAGCTCA CCTGCCCGGT GATGGCCCTG
34561 GCCGGCGACC GTGACCCGAA GGCGCCGCTG AACGAGGTGG CCGAGTGGCG TCGGCACACC
34621 AGCGGGCCGT TCTGCCTCCG GCGTACTCC GGCGGCCACT TCTACCTCAA CGACAGTGG
34681 CACGAGATCT GCAACGACAT CTCCGACCAC CTGCTCGTCA CCGCGGCGC GCCCGATGCC
34741 CGCGTCGTGC AGCCCCGAC CAGCCTTATC GAAGGAGCGG CGAAGAGATG GCAGAACCCA
50 34801 CGGTGACCGA CGACCTGACG GGGGCCCTCA CGCAGCCCC GCTGGGCCG ACCGTCCGCG
34861 CGGTGGCCGA CCGTGAATC GGCACCCACC TCCTGGAGAC CCGCGGCATC CACTGGATCC
34921 ACGCCGCGAA CCGCGACCCG TACGCCACCG TGCTGCGCGG CCAGGCGGAC GACCCGTATC
34981 CCGCGTACGA GCGGGTGCGT GCCCGCGCG CGCTCTCCTT CAGCCCGACG GGCAGCTGGG
35041 TCACCGCCGA TCACGCCCTG GCGGCGAGCA TCCTCTGCTC GACGGACTTC GGGGTCTCCG
55 35101 GCGCCGACGG CGTCCCGGTG CCGCAGCAG TCCTCTCGTA CGGGGAGGGC TGTCCGCTGG
35161 AGCGCGAGCA GGTGCTGCCG GCGGCCGGT ACGTGCCGGA GGGCGGGCAG CGTGCCGTGG
35221 TCGAGGGGAT CCACCGGGAG ACGCTGGAG GTCTCGCGCC GGACCCGTCG GCGTCGTACG
35281 CCTTCGAGCT GCTGGGCGGT TTCGTCCGCC CGGCGGTGAC GGCCGCTGCC GCCCGCGTGC
35341 TGGGTGTTCC CGCGGACCG CGCGCGGACT TCGCGGATCT GCTGGAGCG CTCCGGCCGC
35401 TGTCCGACAG CCTGCTGGCC CCGCAGTCCC TGCGGACGGT ACGGGCGGCG GACGGCGCGC


```

35461 TGGCCGAGCT CACGGCGCTG CTCGCCGATT CGGACGACTC CCCCAGGGGCC CTGCTGTCGG
35521 CGCTCGGGGT CACCGCAGCC GTCCAGCTCA CCGGGAACGC GGTGCTCGCG CTCCTCGCGC
35581 ATCCCGAGCA GTGGCGGGAG CTGTGCGACC GGCCCGGGCT CGCGGCGGCC GCGGTGGAGG
35641 AGACCCTCCG CTACGACCCG CCGGTGCAGC TCGACGCCCG GGTGGTCCGC GGGGAGACGG
5 35701 AGCTGGCGGG CCGGCGGCTG CCGGCCGGGG CGCATGTCGT CGTCCTGACC GCCGCGACCG
35761 GCCGGGACCC GGAGGTCTTC ACGGACCCGG AGCGCTTCGA CCTCGCGCGC CCCGACGCCG
35821 CCGCGCACCT CGCGCTGCAC CCCGCCGGTC CGTACGGCCC GGTGGCGTCC CTGGTCCGGC
35881 TTCAGGCGGA GGTGCGCTG CCGACCCTGG CCGGGCGTTT CCCCAGGGCTG CCGCAGGCGG
35941 GGGACGTGCT CCGCCCCCGC CGCGCGCCTG TCGGCCGCGG GCCGCTGAGC GTCCCGGTCA
10 36001 GCAGCTCCTG AGACACCGGG GCCCCGGTCC GCCCGGCCCC CCTTCGGACG GACCGGACGG
36061 CTCGGACCAC GGGGACGGCT CAGACCGTCC CGTGTGTCCC CGTCCGGCTC CCGTCCGCCC
36121 CATCCCGCCC CTCCACCGGC AAGGAAGGAC ACGACGCCAT GCGCGTCTCT CTGACCTCGT
36181 TCGCACATCA CACGCACTAC TACGGCCTGG TGCCCTTGGC CTGGGCGCTG CTCGCCGCGG
36241 GGCACGAGGT GCGGGTCGCC AGCCAGCCCG CGCTACGGA CACCATCACC GGGTCCGGGC
15 36301 TCGCCGCGGT GCCGGTCGGC ACCGACCACC TCATCCACGA GTACCGGGTG CCGATGGCGG
36361 GCGAGCCGCG CCCGAACCAT CCGGCGATCG CCTTCGACGA GGCCCGTCCC GAGCCGCTGG
36421 ACTGGGACCA CGCCCTCGGC ATCGAGGCGA TCCTCGCCCC GTACTTCTAT CTGCTCGCCA
36481 ACAACGACTC GATGGTCGAC GACCTCGTCG ACTTCGCCCG GTCCTGGCAG CCGGACCTGG
36541 TGCTGTGGGA GCCGACGACC TACGCGGGCG CCGTCGCCGC CCAGGTCACC GGTGCCGCGC
20 36601 ACGCCCGGGT CCTGTGGGGG CCCGACGTGA TGGGACGCGC CCGCCGCAAG TTCGTGCGCG
36661 TCGGGGACCG GCAGCCGCCC GAGCACCAGG AGGACCCAC CGCGGAGTGG CTGACGTGGA
36721 CGCTCGACCG GTACGGCGCC TCCTTCGAAG AGGAGCTGCT CACCGGCCAG TTCACGATCG
36781 ACCCGACCCC GCCGAGCCTG CGCCTCGACA CGGGCCTGCC GACCGTCGGG ATGCGTTATG
36841 TTCCGTACAA CCGCACGTCG GTCGTGCCGG ACTGGCTGAG TGAGCCGCCC GCGCGGCCCC
25 36901 GGGTCTGCCT GACCCTCGGC GTCTCCGCGC GTGAGGTCCT CGGCGGCGAC GCGGTCTCGC
36961 AGGGCGACAT CCTGGAGGCG CTCGCCGACC TCGACATCGA GCTCGTCGCC ACGCTCGACG
37021 CGAGTCAGCG CGCCGAGATC CGCAACTACC CGAAGCACAC CCGGTTACG GACTTCGTGC
37081 CGATGCACGC GCTCCTGCCG AGCTGCTCGG CGATCATCCA CCACGGCGGG GCGGGCACCT
37141 ACGCGACCGC CGTGATCAAC GCGGTGCCGC AGGTATGCT CGCCGAGCTG TGGGACGCGC
30 37201 CGGTCAAGGC GCGGGCCGTC GCCGAGCAGG GGGCGGGGTT CTTCTGCGG CCGGCCGAGC
37261 TCACGCCGCA GGCGTGCGG GACGCCGTCG TCCGCATCCT CGACGACCCC TCGGTCGCCA
37321 CCGCCGCGCA CCGGCTGCGC GAGGAGACCT TCGGCGACCC CACCCCGGCC GGGATCGTCC
37381 CCGAGCTGGA GCGGCTCGCC GCGCAGCACC GCCGCCCGCC GGCCGACGCC CCGCACTGAG
37441 CCGCACCCCT CGCCCCAGGC CTCACCCCTG TATCTGCGCC GGGGGACGCC CCCGGCCAC
35 37501 CCTCCGAAAG ACCGAAAGCA GGAGACCGT GTACGAAGTC GACCACGCCG ACGTCTACGA
37561 CCTCTTCTAC CTGGGTCGCG GCAAGGACTA CGCCGCCGAG GCCTCCGACA TCGCCGACCT
37621 GGTGCGCTCC CGTACCCCCG AGGCCTCCTC GTCCTGGAC GTGGCCTGCG GTACGGGCAC
37681 GCATCTGGAG CACTTCACCA AGGAGTTCGG CGACACCGCC GGCCTGGAGC TGTCGAGGA
37741 CATGCTCACC CACGCCCCGA AGCGGCTGCC CGACGCCACG CTCCACCAGG GCGACATGCG
40 37801 GGAATTCCGG CTCGGCCGGA AGTTCTCCGC CGTGGTCAGC ATGTTACAGT CCGTCGGCTA
37861 CCTGAAGACG ACCGAGGAAC TCGGCGCGGC CGTCGCCTCG TTCGCGGAGC ACCTGGAGCC
37921 CCGTGGCGTC GTCGTGTCG AGCCGTGGTG GTTCCCGGAG ACCTTCGCCG ACGGCTGGGT
37981 CAGCGCCGAC GTCGTCCGCC GTGACGGGCG CACCGTGGCC CGTGTCTCGC ACTCGGTGCG
45 38041 GGAGGGGAAC GCGACGCGCA TGGAGGTCCA CTTACCCGTG GCCGACCCGG GCAAGGGCGT
38101 GCGGCACTTC TCCGACGTCC ATCTCATCAC CCTGTTCCAC CAGGCCGAGT ACGAGGCCGC
38161 GTTACAGGCC GCCGGGCTGC GCGTCGAGTA CCTGGAGGGC GGCCCGTCGG GCCGTGGCCT
38221 CTTGTCGGC GTCCCCGCT GAGCACCGCC CAAGACCCCC CGGGGCGGGA CGTCCCGGGT
38281 GCACCAAGCA AAGAGAGAGA AACGAACCGT GACAGGTAAG ACCCGAATAC CGCGTGTCCG
38341 CCGCGGCCGC ACCACGCCCA GGGCCTTCAC CCTGGCCGTC GTCGGCACCC TGCTGGCGGG
50 38401 CACCACCGTG GCGGCCGCG CTCCCGCGC CGCCGACACG GCCAATGTT AGTACACGAG
38461 CCGGGCGGCG GAGCTCGTCG CCCAGATGAC GTCGACGAG AAGATC (SEQ ID NO:19)

```

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the narbonolide PKS of *Streptomyces venezuelae* is shown herein merely to illustrate a

preferred embodiment of the invention, and the invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant
5 loss of a desired activity. The present invention includes such polypeptides with alternate amino acid sequences, and the amino acid sequences shown merely illustrate preferred embodiments of the invention.

The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and
10 methods provided thereby, the following description of the various regions of the narbonolide PKS and corresponding coding sequences is provided.

The loading module of the narbonolide PKS contains an inactivated KS domain, an AT domain, and an ACP domain. The AT domain of the loading module binds propionyl CoA. Sequence analysis of the DNA encoding the KS domain indicates that this domain is
15 enzymatically inactivated, as a critical cysteine residue in the motif TVDACSSSL, which is highly conserved among KS domains, is replaced by a glutamine and so is referred to as a KS^Q domain. Such inactivated KS domains are also found in the PKS enzymes that synthesize the 16-membered macrolides carbomycin, spiromycin, tylosin, and niddamycin. While the KS domain is inactive for its usual function in extender modules, it is believed to
20 serve as a decarboxylase in the loading module.

The present invention provides recombinant DNA compounds that encode the loading module of the narbonolide PKS and useful portions thereof. These recombinant DNA compounds are useful in the construction of PKS coding sequences that encode all or a portion of the narbonolide PKS and in the construction of hybrid PKS encoding DNA
25 compounds of the invention, as described in the section concerning hybrid PKSs below. To facilitate description of the invention, reference to a PKS, protein, module, or domain herein can also refer to DNA compounds comprising coding sequences therefor and *vice versa*. Also, reference to a heterologous PKS refers to a PKS or DNA compounds comprising coding sequences therefor from an organism other than *Streptomyces venezuelae*. In addition,
30 reference to a PKS or its coding sequence includes reference to any portion thereof.

The present invention provides recombinant DNA compounds that encode one or more of the domains of each of the six extender modules (modules 1 - 6, inclusive) of the narbonolide PKS. Modules 1 and 5 of the narbonolide PKS are functionally similar. Each of

these extender modules contains a KS domain, an AT domain specific for methylmalonyl CoA, a KR domain, and an ACP domain. Module 2 of the narbonolide PKS contains a KS domain, an AT domain specific for malonyl CoA, a KR domain, a DH domain, and an ACP domain. Module 3 differs from extender modules 1 and 5 only in that it contains an inactive ketoreductase domain. Module 4 of the narbonolide PKS contains a KS domain, an AT domain specific for methylmalonyl CoA, a KR domain, a DH domain, an ER domain, and an ACP domain. Module 6 of the narbonolide PKS contains a KS domain, an AT domain specific for methylmalonyl CoA, and an ACP domain. The approximate boundaries of these "domains" is shown in Table 1.

10 In one important embodiment, the invention provides a recombinant narbonolide PKS that can be used to express only narbonolide (as opposed to the mixture of narbonolide and 10-deoxymethynolide that would otherwise be produced) in recombinant host cells. This recombinant narbonolide PKS results from a fusion of the coding sequences of the *picAIII* and *picAIV* genes so that extender modules 5 and 6 are present on a single protein. This
15 recombinant PKS can be constructed on the *Streptomyces venezuelae* or *S. narbonensis* chromosome by homologous recombination. Alternatively, the recombinant PKS can be constructed on an expression vector and introduced into a heterologous host cell. This recombinant PKS is preferred for the expression of narbonolide and its glycosylated and/or hydroxylated derivatives, because a lesser amount or no 10-deoxymethynolide is produced
20 from the recombinant PKS as compared to the native PKS. In a related embodiment, the invention provides a recombinant narbonolide PKS in which the *picAIV* gene has been rendered inactive by an insertion, deletion, or replacement. This recombinant PKS of the invention is useful in the production of 10-deoxymethynolide and its derivatives without production of narbonolide.

25 In similar fashion, the invention provides recombinant narbonolide PKS in which any of the domains of the native PKS have been deleted or rendered inactive to make the corresponding narbonolide or 10-deoxymethynolide derivative. Thus, the invention also provides recombinant narbonolide PKS genes that differ from the narbonolide PKS gene by one or more deletions. The deletions can encompass one or more modules and/or can be
30 limited to a partial deletion within one or more modules. When a deletion encompasses an entire module, the resulting narbonolide derivative is at least two carbons shorter than the polyketide produced from the PKS encoded by the gene from which deleted PKS gene and corresponding polyketide were derived. When a deletion is within a module, the deletion

typically encompasses a KR, DH, or ER domain, or both DH and ER domains, or both KR and DH domains, or all three KR, DH, and ER domains.

This aspect of the invention is illustrated in Figure 4, parts B and C, which shows how a vector of the invention, plasmid pKOS039-16 (not shown), was used to delete or "knock out" the *picA1* gene from the *Streptomyces venezuelae* chromosome. Plasmid pKOS039-16 comprises two segments (shown as cross-hatched boxes in Figure 4, part B) of DNA flanking the *picA1* gene and isolated from cosmid pKOS023-27 (shown as a linear segment in the Figure) of the invention. When plasmid pKOS039-16 was used to transform *S. venezuelae* and a double crossover homologous recombination event occurred, the *picA1* gene was deleted. The resulting host cell, designated K039-03 in the Figure, does not produce picromycin unless a functional *picA1* gene is introduced.

This *Streptomyces venezuelae* K039-03 host cell and corresponding host cells of the invention are especially useful for the production of polyketides produced from hybrid PKS or narbonolide PKS derivatives. Especially preferred for production in this host cell are narbonolide derivatives produced by PKS enzymes that differ from the narbonolide PKS only in the loading module and/or extender modules 1 and/or 2. These are especially preferred, because one need only introduce into the host cell the modified *picA1* gene or other corresponding gene to produce the desired PKS and corresponding polyketide. These host cells are also preferred for desosaminylating polyketides in accordance with the method of the invention in which a polyketide is provided to an *S. venezuelae* cell and desosaminylated by the endogenous desosamine biosynthesis and desosaminyl transferase gene products.

The recombinant DNA compounds of the invention that encode each of the domains of each of the modules of the narbonolide PKS are also useful in the construction of expression vectors for the heterologous expression of the narbonolide PKS and for the construction of hybrid PKS expression vectors, as described further below.

Section II: The Genes for Desosamine Biosynthesis and Transfer and for Beta-glucosidase

Narbonolide and 10-deoxymethynolide are desosaminylated in *Streptomyces venezuelae* and *S. narbonensis* to yield narbomycin and YC-17, respectively. This conversion requires the biosynthesis of desosamine and the transfer of the desosamine to the substrate polyketides by the enzyme desosaminyl transferase. Like other *Streptomyces*, *S. venezuelae* and *S. narbonensis* produce glucose and a glucosyl transferase enzyme that glucosylates desosamine at the 2' position. However, *S. venezuelae* and *S. narbonensis* also

produce a beta-glucosidase, which removes the glucose residue from the desosamine. The present invention provides recombinant DNA compounds and expression vectors for each of the desosamine biosynthesis enzymes, desosaminyl transferase, and beta-glucosidase.

As noted above, cosmid pKOS023-27 contains three ORFs that encode proteins
5 involved in desosamine biosynthesis and transfer. The first ORF is from the *picCII* gene, also known as *desVIII*, a homologue of *eryCII*, believed to encode a 4-keto-6-deoxyglucose isomerase. The second ORF is from the *picCIII* gene, also known as *desVII*, a homologue of *eryCIII*, which encodes a desosaminyl transferase. The third ORF is from the *picCVI* gene, also known as *desVI*, a homologue of *eryCVI*, which encodes a 3-amino dimethyltransferase.

10 The three genes above and the remaining desosamine biosynthetic genes can be isolated from cosmid pKOS023-26, which was deposited with the American Type Culture Collection on 20 Aug 1998 under the Budapest Treaty and is available under the accession number ATCC 203141. Figure 3 shows a restriction site and function map of cosmid pKOS023-26. This cosmid contains a region of overlap with cosmid pKOS023-27
15 representing nucleotides 14252 to nucleotides 38506 of pKOS023-27.

The remaining desosamine biosynthesis genes on cosmid pKOS023-26 include the following genes. ORF11, also known as *desR*, encodes beta-glucosidase and has no ery gene homologue. The *picCI* gene, also known as *desV*, is a homologue of *eryCI*. ORF14, also known as *desIV*, has no known ery gene homologue and encodes an NDP glucose 4,6-
20 dehydratase. ORF13, also known as *desIII*, has no known ery gene homologue and encodes an NDP glucose synthase. The *picCV* gene, also known as *desII*, a homologue of *eryCV* is required for desosamine biosynthesis. The *picCIV* gene also known as *desI*, is a homologue of *eryCIV*, and its product is believed to be a 3,4-dehydratase. Other ORFs on cosmid pKOS023-26 include ORF12, believed to be a regulatory gene; ORF15, which encodes an S-
25 adenosyl methionine synthase; and ORF16, which is a homolog of the *M. tuberculosis cbhK* gene. Cosmid pKOS023-26 also encodes the *picK* gene, which encodes the cytochrome P450 hydroxylase that hydroxylates the C12 of narbomycin and the C10 and C12 positions of YC-17. This gene is described in more detail in the following section.

Below, the amino acid sequences or partial amino acid sequences of the gene products
30 of the desosamine biosynthesis and transfer and beta-glucosidase genes are shown. These amino acid sequences are followed by the DNA sequences that encode them.

Amino acid sequence of PICCI (*desV*) (SEQ ID NO:6)

1 VSSRAETPRV PFLDLKAAEY ELRAETDAAI ARVLDSGRYL LGPELEGFEA EFAAYCETDH
 61 AVGVNSGMDA LQLALRGLGI GPGDEVIVPS HTYIASWLAV SATGATPVVPV EPHEDHPTLD
 121 PLLVEKAITP RTRALLPVHL YGHPADMDAL RELADRHGLH IVEDAAQAHG ARYRGRRIGA
 5 181 GSSVAAFSPY PGKNLGCFGD GGAVVTGDPE LAERLRMLRN YGSRQKYSHE TKGTSNRLDE
 241 MQAAVLRIRL XHLD SWNGRR SALAAEYLSG LAGLPGIGLP VTAPDTPVW HLFTVTRERR
 301 DELRSHLDAR GIDTLTHYPV PVHLSPAYAG EAPPEGSLPR AESFARQVLS LPIGPHLERP
 361 QALRVIDAVR EWAERVDQA (SEQ ID NO:6)

10 Amino acid sequence of 3-keto-6-deoxyglucose isomerase, PICCII (*desVIII*) (SEQ ID NO:7)

1 VADRELGTHL LETRGIHWIH AANGDPYATV LRGQADDPYP AYERVRARGA LSFSPGTSWV
 61 TADHALAASI LCSTDFGVSG ADGVVPVQOV LSYGEGCPLE REQVLPAAGD VPEGGQRAVV
 121 EGIHRETLEG LAPDPSASYA FELLGGFVRP AVTAAAAAVL GVPADRRADF ADLLERLRPL
 181 SDSLLAPQSL RTVRAADGAL AELTALLADS DDSFGALLSA LGVTAAVQLT GNAVLALLAH
 15 241 PEQWRELCDR PGLAAA AVEE TLRYDPPVQL DARVVRGETE LAGRRLPAGA HVVVLTAATG
 301 RDPEVFTDPE RFDLARPDAA AHLALHPAGP YGPVASLVRL QAEVALRTLA GRFPGLRQAG
 361 DVLRRRAPV GRGPLSVVPS SS (SEQ ID NO:7)

Amino acid sequence of desosaminyl transferase, PICCIII (*desVII*) (SEQ ID NO:8)

20 1 MRVLLTSFAH HTHYYGLVPL AWALLAAGHE VRVASQPALT DTITGSGLAA VPVGTDLHIH
 61 EYVRMAGEP RPNHPAIAFD EARPEPLDWD HALGIEAILA PYFYLLANND SMVDDLVDFA
 121 RSWQPDVLW EPTTYAGAVA AQVTGAHAR VLWGPDMVMS ARRKFVALRD RQPPEHREDP
 181 TAEWLTWTLT RYGASFEEEL LTGQFTIDPT PPSRLRLDTGL PTVGMRYVPY NGTSVVPDWL
 241 SEPPARPRVC LTLGVSAREV LGGDGVSQGD ILEALADLDI ELVATLDASQ RAEIRNYPKH
 25 301 TRFTDFVPMH ALLPSCSAII HHGGAGTYAT AVINAVPQVM LAELWDAPVK ARAVAEQGAG
 361 FFLPPAELTP QAVRDAVVRI LDDPSVATAA HRLREETFGD PTPAGIVPEL ERLAAQHRRP
 421 PADARH (SEQ ID NO:8)

Partial amino acid sequence of aminotransferase-dehydrase, PICCIV (*desI*) (SEQ ID NO:9)

30 1 VKSALSDIAF FGGPAAFDQP LLVGRPNRID RARLYERLDR ALDSQWLSNG GPLVREFEER
 61 VAGLAGVRHA VATCNATAGL QLLAHAAGLT GEVIMPSMTF AATPHALRWI GLTPVFADID
 121 PDTGNLDPDQ VAAAVTPRTS AVVGVLWGR PCAADQLRKV ADEHGLRLYF DAAHALGCAV
 181 DGRPAGSLGD AEVFSFHATK AVNAFEGGAV VTDDADLAAR IRALHNFGFD LPGGSPAGGT
 241 NAKMSEAAAA MGLTSLDAFP EVIDRNRNRH AXIREHLADL PGVLVADHDR HGLNNHQYVI
 35 301 VEIDEATTGI HRDLVMEVLK AEGVHTRAYF S (SEQ ID NO:9)

Amino acid sequence of PICCV (*desII*) (SEQ ID NO:10)

1 MTAPALSATA PAERCAHPGA DLGAHVAVG QTLAAGGLVP PDEAGTTARH LVRLAVRYGN
 61 SPFTPLEEAR HDLGVD RDAF RRLALFGQV PELRTAVETG PAGAYWKNTL LPLEQRGVFD
 40 121 AALARKPVFP YSVGLYPGPT CMFRCHFCVR VTGARYDPSA LDAGNAMERS VIDEIPAGNP
 181 SAMYFSGGLE PLTNPGLGSL AAHATDHGLR PTVYTNSFAL TERTLERQPG LWGLHAIRTS
 241 LYGLNDEEYE QTTGKKA AFR RVRENLR RFQ QLRAERESPI NLGFAYIVLP GRASRLDLV
 301 DFIADLNDAG QGRTIDFVNI REDYSGRDDG KLPQEERAEL QEALNAFEER VRERTPGLHI
 361 DYGYALNSLR TGADAELLRI KPATMRPTAH PQVAVQVDLL GDVLYREAG FPDLDGATRY
 45 421 IAGRVTPDTS LTEVVRDFVE RGGEVA AVDG DEYFMDGFDQ VVTARLNQLE RDAADGWEEA
 481 RGFLR (SEQ ID NO:10)

Amino acid sequence of 3-amino dimethyl transferase, PICCVI (*desVI*) (SEQ ID NO:11)

1 VYEVHADVY DLFYLGRGKD YAAEASDIAD LVRSTPEAS SLLDVACGTG THLEHFTKEF
 50 61 GDTAGLELSE DMLTHARKRL PDATLHQGDM RDFRLGRKFS AVVSMFSSVG YLKTTEELGA
 121 AVASFAEHLE PGGVVVVEPW WFPETFADGW VSADVVRDGD RTVARVSHSV REGNATRMEV
 181 HFTVADPGKG VRHFS DVHLI TLFHQAEYEA AFTAAGLRVE YLEGGPSGRG LFVGVPA (SEQ
 ID NO:11)

Partial amino acid sequence of beta-glucosidase, ORF11 (*desR*) (SEQ ID NO:12)

1 MTLDEKISFV HWALDPDRQN VGYLPGVPRL GIPELRAADG PNGIRLVGQT ATALPAPVAL
 61 ASTFDDTMAD SYGKVMGRDG RALNQDMVLG PMMNIRVPH GGRNYETFSE DPLVSSRTAV
 5 121 AQIKGIQAG LMTTAKHFAA NNQENNRFSV NANVDEQTLR EIEFFAFEAS SKAGAGSFC
 181 AYNGLNKPS CGNDELLNV LRTQWGFQGW VMSDWLATPG TDAITKGLDQ EMGVELPGDV
 241 PKGEPSPPAK FFGEALKTAV LNGLTVPEAAV TRSAERIVGQ MEKFGLLLAT PAPRPERDKA
 301 GAQAVSRKVA ENGAVLLRNE GQALPLAGDA GKSIAVIGPT AVDPKVTGLG SAHVVPDSAA
 361 APLDTIKARA GAGATVTYET GEETFGTQIP AGNLSPAFNQ GHQLEPGKAG ALYDGTTLTVP
 10 421 ADGEYRIAVR ATGGYATVQL GSHTIEAGQV YGKVSSPLLK LTKGTHKLT I SGFAMSATPL
 481 SLELGWVTPA AADATIAKAV ESARKARTAV VFAYDDGTEG VDRPNLSLPG TQDKLISAVA
 541 DANPNTIVVL NTGSSVLMPW LSKTRAVLDM WYPGQAGAEA TAALLYGDVN PSGKLTQSFP
 601 AAENQHAVAG DPTSYPGVDN QQTYREGIHV GYRWFDKENV KPLFPFGHGL SYTSFTQSAP
 661 TVVRTSTGGL KVTVTVRNSG KRAGQEVVQA YLGASPNVTA PQAKKKLVGY TKVSLAAGEA
 15 721 KTVTNNVDRR QLQFWDAATD NWKTGTGNRL LQTGSSSADL RGSATVNVW (SEQ ID NO:12)

Amino acid sequence of transcriptional activator, ORF12 (regulatory) (SEQ ID NO:13)

1 MNLVERDGEI AHLRAVLDA AAGDGTLLLV SGPAGSGKTE LLRSLRRLAA ERETPVWSVR
 61 ALPGDRDIPL GVLCQLLRS EQHGADTSV RDLLDAASRR AGTSPPPTR RSASTRHTAC
 20 121 TTGCSPSPAG TPFLVAVDDL THADTASLRF LLYCAAHHDQ GGIGFVMTER ASQRAGYRVF
 181 RAELLRQPHC RNMWLSGLPP SGVRQLLAHY YGPEAAERRA PAYHATTGGN PLLLRALTQD
 241 RQASHTTLGA AGGDEPVHGD AFAQAVLDCL HRSAGTLET ARWLAVLEQS DPLLVERLTG
 301 TTAAVERHI QELAAIGLLD EDGTLGQPAI REAALQDLPA GERTELHRA AEQLHRDGAD
 361 EDTVARHLLV GGAPDAPWAL PLLERGAQQA LFDDRLDDAF RILEFAVRSS TDNTQLARLA
 25 421 PHLVAASWRM NPHMTTRALA LFDRLSGEL PPSHPVMALI RCLVWYGRLP EAADALSRLR
 481 PSSDNDALEL SLTRMWLAAL CPPLLESPLA TPEPERGPVP VRLAPRTTAL QAQAGVFQRG
 541 PDNASVAQAE QILQGCRLSE ETYEALETAL LVLVHADRLD RALFWSALL AEAVERRSLG
 601 WEAVFAATRA MIAIRCGDLP TARERAELAL SHAAPESWGL AVGMPLSALL LACTEAGEYE
 661 QAERVLQPV PDAMFDSRHG MEYMHARGRY WLAXGRLHAA LGEFMLCGEI LGSWNLDQPS
 30 721 IVPWRTSAAE VYLRLGNRQK ARALAEQLA LVRPGRSRT GLTLRVLAAA VDGQQAERLH
 781 AEAVDMLHDS GDRLEHARAL AGMSRHQQAQ GDNYRARMTA RLAGDMAWAC GAYPLAEEIV
 841 PGRGGRRKA VSTELELPGG PDVGLLSEAE RRVAALAARG LTNRQIARRL CVTASTVEQH
 901 LTRVYRKLNV TRRADLPISL AQDKSVTA (SEQ ID NO:13)

 35 Amino acid sequence of dNDP-glucose synthase (glucose-1-phosphate thymidyl transferase),
 ORF13 (*desIII*) (SEQ ID NO:14)

1 MKGIVLAGGS GTRLHPATSV ISKQILPVYN KPMIYPLSV LMLGGIREIQ IISTPQHIEL
 61 FQSLGNGRH LGIELDYAVQ KEPAGIADAL LVGAEHIGDD TCALILGDNI FHGPGLYTLL
 121 RDSIARLDGC VLFYGPVKDP ERYGVAEVDA TGRITDLVEK PVKPRSNLAV TGLYLYDNDV
 40 181 VDIKNIRPS PRGELEITDV NRVYLERGRA ELVNLGRGFA WLDGTGTHDSL LRAAQYVQVL
 241 EERQGVWIA LEEIAFRMGF IDAEACHGLG EGLSRTEYGS YLMEIAGREG AP (SEQ ID
 NO:14)

Amino acid sequence of dNDP-glucose 4,6-dehydratase, ORF14 (*desIV*) (SEQ ID NO:15)

1 VRLLVTGGAG FIGSHFVRQL LAGAYPDVPA DEVIVLDSLT YAGNRANLAP VDADPRLRFV
 61 HGDIRDAGLL ARELRGVDAI VHFAAESHVD RSIAGASVFT ETNVQGTQTL LQCAVDAGVG
 121 RVVHVSTDEV YGSIDSGSWT ESSPLEPNP YAASKAGSDL VARAYHRTYG LDVRITRCN
 181 NYGPYQHPEK LIPLFVTNLL DGGTLPLYGD GANVREWVHT DDHCRGIALV LAGGRAGEIY
 241 HIGGLELTN RELTGILLDS LGADWSSVRK VADRKGHDLR YSLDGGKIER ELGYRPQVSF
 50 301 ADGLARTVRW YRENRGWEP LKATAPQLPA TAVEVSA (SEQ ID NO:15)

 Partial amino acid sequence of S-adenosylmethionine synthase, ORF15 (SAM synthase)
 (SEQ ID NO:16)

1 IGYDSSKKGF DGASCGSVS IGSQSPDIAQ GVDYAYEKRV EGASQRDEGD ELDKQAGDQ

5 61 GLMFGYASDE TPELMPLPIH LAHRLSRRLT EVRKNGTIPY LRPDGKTQVT IEYDGDRAVR
 121 LDTVVSQH ASDIDLESLL APDVRKFVVE HVLAQLVEDG IKLDTDGYRL LVNPTGRFEI
 181 GGPMGDAGLT GRKIIDTYG GMARHGGGAF SGKDPKVD R SAAYAMRWVA KNVVAAGLAS
 241 RCEVQVAYAI GKAEPVGLFV ETEFGTHKIET EKIEAIGEV FDLRPAAIIR DLDLLRPIYS
 301 QTAAYGHFGR ELPDFTWERT DRVDALKKAA GL (SEQ ID NO:16)

Partial amino acid sequence of ORF16 (homologous to *M. tuberculosis* cbhK) (SEQ ID NO:17)

10 1 MRIAVTGSIA TDHLMTFPGR FAEQILPDQL AHVSLSFLVD TLDIRHGGVA ANIAYGLGLL
 61 GRRPVLVGAV GKDFDGYGQL LRAAGVDTDS VRVSDRQHTA RFMCTTDEDG NQLASFYAGA
 121 MAEARDIDLG ETAGRPGGID LVLVGADDOPE AMVRHTRVCR ELGLRRAADP SQQLARLEGD
 181 SVRELVDGAE LLFTNAYERA LLLSKTGWTE QEVLARVGTW ITTLGAKGCR (SEQ ID NO:17)

15 While not all of the insert DNA of cosmid pKOS023-26 has been sequenced, five
 large contigs shown of Figure 3 have been assembled and provide sufficient sequence
 information to manipulate the genes therein in accordance with the methods of the invention.
 The sequences of each of these five contigs are shown below.

20 Contig 001 from cosmid pKOS023-26 contains 2401 nucleotides, the first 100 bases
 of which correspond to 100 bases of the insert sequence of cosmid pKOS023-27. Nucleotides
 80 - 2389 constitute ORF11, which encodes 1 beta glucosidase. (SEQ ID NO:20)

1 CGTGGCGGCC GCCGCTCCCG GCGCCGCCGA CACGGCCAAT GTTCAGTACA CGAGCCGGGC
 61 GGCGGAGCTC GTCGCCCAGA TGACGCTCGA CGAGAAGATC AGCTTCGTCC ACTGGGCGCT
 121 GGACCCCGAC CGGCAGAACG TCGGCTACCT TCCCGGCGTG CCGCGTCTGG GCATCCCGGA
 181 GCTGCGTGCC GCGACGGCC CGAACGGCAT CCGCCTGGTG GGGCAGACCG CCACCGCGCT
 25 241 GCGCGCGCCG GTCGCCCTGG CCAGCACCTT CGACGACACC ATGGCCGACA GCTACGGCAA
 301 GGTGATGGGC CGCGACGGTC GCGCGCTCAA CCAGGACATG GTCCTGGGCC CGATGATGAA
 361 CAACATCCGG GTGCCGACG GCGGCCGGA CTACGAGACC TTCAGCGAGG ACCCCCTGGT
 421 CTCCTCGCGC ACCGCGGTCG CCCAGATCAA GGGCATCCAG GGTGCGGGTC TGATGACCAC
 481 GGCCAAGCAC TTCGCGGCCA ACAACCAGGA GAACAACCGC TTCTCCGTGA ACGCCAATGT
 30 541 CGACGAGCAG ACGCTCCGCG AGATCGAGTT CCCGGCGTTC GAGGCGTCCT CCAAGGCCGG
 601 CGCGGGCTCC TTCATGTGTG CCTACAACGG CCTCAACGGG AAGCCGTCCT GCGGCAACGA
 661 CGAGCTCCTC AACAACGTGC TGCGCACGCA GTGGGGCTTC CAGGGCTGGG TGATGTCCGA
 721 CTGGCTCGCC ACCCCGGGCA CCGACGCCAT CACCAAGGGC CTCGACCAGG AGATGGGCGT
 781 CGAGCTCCCC GGCGACGTCC CGAAGGGCGA GCCCTCGCCG CCGGCCAAGT TCTTCGGCGA
 35 841 GGCGCTGAAG ACGGCCGTCC TGAACGGCAC GGTCCCCGAG GCGGCCGTGA CGCGGTGCGG
 901 GGAGCGGATC GTCGGCCAGA TGGAGAAGTT CGGTCTGCTC CTCGCCACTC CGGCGCCGCG
 961 GCGGAGCGC GACAAGGCGG GTGCCAGGC GGTGTCCCGC AAGGTCGCCG AGAACGGCGC
 1021 GGTGCTCCTG CGCAACGAGG GCCAGGCCCT GCGGCTCGCC GGTGACGCCG GCAAGAGCAT
 1081 CGCGGTCATC GGCCCGACGG CCGTCGACCC CAAGGTCACC GGCTGGGCA GCGCCACGT
 40 1141 CGTCCCGGAC TCGGCGGCGG CGCCACTCGA CACCATCAAG GCCCGCGCGG GTGCGGGTGC
 1201 GACGGTGACG TACGAGACGG GTGAGGAGAC CTTGCGGACG CAGATCCCGG CGGGGAACCT
 1261 CAGCCCGGCG TTCAACCAGG GCCACCAGCT CGAGCCGGGC AAGGCGGGG CGCTGTACGA
 1321 CGGCACGCTG ACCGTGCCCC CCGACGGCGA GTACCGCATC GCGGTCCGTG CCACCGGTGG
 1381 TTACGCCACG GTGCAGCTCG GCAGCCACAC CATCGAGGCC GGTGAGGTCT ACGGCAAGGT
 45 1441 GAGCAGCCCC CTCCTCAAGC TGACCAAGGG CACGCACAAG CTCACGATCT CGGGCTTCGC
 1501 GATGAGTGCC ACCCGCTCT CCCTGGAGCT GGGCTGGGTN ACGCCGGCGG CGGCCGACGC
 1561 GACGATCGCG AAGGCCGTGG AGTCGGCGCG GAAGGCCCGT ACGGCGGTCT TCTTCGCTA
 1621 CGACGACGGC ACCGAGGGCG TCGACCGTCC GAACCTGTCT CTGCCGGGTA CGCAGGACAA
 1681 GCTGATCTCG GCTGTGCGGG ACGCCAACCC GAACACGATC GTGGTCCTCA ACACCGGTT
 50 1741 GTCGGTGCTG ATGCCGTGGC TGTCCAAGAC CCGCGCGGTC CTGGACATGT GGTACCCGGG
 1801 CCAGGCGGGC GCGGAGGCGA CCGCGCGCT GCTCTACGGT GACGTCAACC CGAGCGGCAA
 1861 GCTCACGCAG AGCTTCCCGG CCGCGGAGAA CCAGCACGCG GTCGCCGGCG ACCCGACCAG

1921 CTACCCGGGC GTCGACAACC AGCAGACGTA CCGCGAGGGC ATCCACGTCG GGTACCGCTG
 1981 GTTCGACAAG GAGAACGTCA AGCCGCTGTT CCCGTTTCGGG CACGGCCTGT CGTACACCTC
 2041 GTTCACGCAG AGCGCCCCGA CCGTCGTGCG TACGTCCACG GGTGGTCTGA AGGTCACGGT
 2101 CACGGTCCGC AACAGCGGGA AGCGCGCCGG CCAGGAGGTC GTCCAGGCGT ACCTCGGTGC
 5 2161 CAGCCCGAAC GTGACGGCTC CGCAGGCGAA GAAGAAGCTC GTGGGCTACA CGAAGGTCTC
 2221 GCTCGCCGCG GGCGAGGCGA AGACGGTGAC GGTGAACGTC GACCGCCGTC AGCTGCAGTT
 2281 CTGGGATGCC GCCACGGACA ACTGGAAGAC GGGAACGGGC AACC GCCTCC TGCAGACCGG
 2341 TTCGTCCTCC GCCGACCTGC GGGGCAGCGC CACGGTCAAC GTCTGGTGAC GTGACGCCGT
 2401 G (SEQ ID NO:20)
 10

Contig 002 from cosmid pKOS023-26 contains 5970 nucleotides and the following
 ORFs: from nucleotide 995 to 1 is an ORF of *picCIV* that encodes a partial sequence of an
 amino transferase-dehydrase; from nucleotides 1356 to 2606 is an ORF of *picK* that encodes
 a cytochrome P450 hydroxylase; and from nucleotides 2739 to 5525 is ORF12, which
 15 encodes a transcriptional activator. (SEQ ID NO:21)

1 GCGGAGAAGT AGGCGCGGGT GTGCACGCCT TCGGCCTTCA GGACCTCCAT GACGAGGTCTG
 61 CGGTGGATGC CGGTGGTGGC CTCGTCGATC TCGACGATCA CGTACTGGTG GTTGTGAGG
 121 CCGTGGCGGT CGTGGTCGGC GACGAGGACG CCGGGGAGGT CCGCGAGGTG CTCGCGGTAG
 181 SCGGCGTGGT TGCGCCGGTT CCGGTGATG ACCTCGGGAA ACGCGTCGAG GGAGGTGAGG
 20 241 CCCATGGCGG CGGCGGCCTC GCTCATCTTG GCGTTGGTCC CGCCGGCGGG GCTGCCGCCG
 301 GGCAGGTCGA AGCCGAAGTT GTGGAGGGCG CGGATCCGGG CGGCGAGGTC GGCGTCGTCTG
 361 GTGACGACGG CGCCGCCCTC GAAGGCGTTG ACGGCCTTGG TGGCGTGGAA GCTGAAGACC
 421 TCGGCGTCGC CGAGGCTGCC GGCGGGCCGG CCGTCGACCG CGCAGCCGAG GGCCTGCGCG
 481 GCGTCGAAGT ACAGCCGCAG GCCGTGCTCG TCGGCGACCT TCCGCAGCTG GTCGGCGGCG
 25 541 CAGGGGCGGC CCCAGAGGTG GACGCCGACG ACGGCCGAGG TGCGGGGTGT GACCGCGGCG
 601 GCCACCTGGT CCGGGTCGAG GTTGCCGGTG TCCGGGTCGA TGTGCGCGAA GACCGGGGTG
 661 AGGCCGATCC AGCGCAGTGC GTGCGGGGTG GCGGCGAACG TCATCGACGG CATGATCACT
 721 TCGCCGGTGA GGCCGGCGGC GTGCGCGAGG AGCTGGAGCC CGGCCGTGGC GTTGCAGGTG
 781 GCCACGGCAT GCCGGACCCC GCGGAGCCCG GCGACGCGCT CCTCGAACTC GCGGACGAGC
 30 841 GGGCCGCCGT TGGACAGCCA CTGGCTGTCG AGGGCCCCGT CGAGCCGCTC GTACAGCCTG
 901 GCGCGGTCTGA TGCGGTTGGG CCGCCCCACG AGGAGCGGCT GGTGAAAGC GCGGGGCGCG
 961 CCGAAGAATG CGAGGTCGGA TAAGGCGCTT TTCACGGATG TTCCCTCCGG GCCACCGTCA
 1021 CGAAATGATT CGCCGATCCG GGAATCCCGA ACGAGGTCGC CGCGCTCCAC CGTGACGTAC
 1081 GACGAGATGG TCGATTGTGG TGGTCGATTT CCGGGGGACT CTAATCCGCG CGGAACGGGA
 35 1141 CCGACAAGAG CACGCTATGC GCTCTCGATG TGCTTCGGAT CACATCCGCC TCCGGGGTAT
 1201 TCCATCGGCG GCCCGAATGT GATGATCCTT GACAGGATCC GGAATCAGC CGAGCCGCCG
 1261 GGAGGGCCGG GGCGCGCTCC GCGGAAGAGT ACGTGTGAGA AGTCCCGTTC CTCTTCCCGT
 1321 TTCCGTTCCG CTTCCGGCCC GGTCTGGAGT TCTCCGTGCG CCGTACCCAG CAGGGAACGA
 1381 CCGCTTCTCC CCCGGTACTC GACCTCGGGG CCCTGGGGCA GGATTTTCGCG GCCGATCCGT
 40 1441 ATCCGACGTA CGCGAGACTG CGTGCCGAGG GTCCGGCCCA CCGGGTGCGC ACCCCGAGG
 1501 GGGACGAGGT GTGGCTGGTC GTCGGCTACG ACCGGGCGCG GGCGGTCTC GCCGATCCCC
 1561 GGTTCAGCAA GGAAGTGGCG AACTCCACGA CTCCCTGAC CGAGGCCGAG GCCGCGCTCA
 1621 ACCACAACAT GCTGGAGTCC GACCCGCCGC GGCACACCCG GCTGCGCAAG CTGGTGGCCC
 1681 GTGAGTTCAC CATGCGCCGG GTCGAGTTGC TGCGGCCCCG GGTCCAGGAG ATCGTCGACG
 45 1741 GGCTCGTGGA CGCCATGCTG GCGGCGCCCG ACGGCCGCGC CGATCTGATG GAGTCCCTGG
 1801 CCTGGCCGCT GCCGATCACC GTGATCTCCG AACTCCTCGG CGTGCCCGAG CCGGACCGCG
 1861 CCGCCTTCCG CGTCTGGACC GACGCCTTCG TCTTCCCGGA CGATCCCGCC CAGGCCCAGA
 1921 CCGCCATGGC CGAGATGAGC GGCTATCTCT CCCGGCTCAT CGACTCCAAG CGCGGGCAGG
 1981 ACGGCGAGGA CCTGCTCAGC GCGCTCGTGC GGACCAGCGA CGAGGACGGC TCCCGGCTGA
 50 2041 CCTCCGAGGA GCTGCTCGGT ATGGCCCACA TCCTGCTCGT CGCGGGGCAC GAGACCACGG
 2101 TCAATCTGAT CGCCAACGGC ATGTACGCGC TGCTCTCGCA CCCCAGCCAG CTGGCCGCCC
 2161 TCGGGGCCGA CATGACGCTC TTGGACGGCG CCGTGGAGGA GATGTTGCGC TACGAGGGCC
 2221 CGGTGGAATC CGCGACCTAC CGCTTCCCGG TCGAGCCCGT CGACCTGGAC GGCACGGTCA
 2281 TCCCGGCCGG TGACACGGTC CTCGTCGTCC TGGCCGACGC CCACCGCACC CCCGAGCGCT

2341 TCCCGGACCC GCACCGCTTC GACATCCGCC GGGACACCGC CGGCCATCTC GCCTTCGGCC
2401 ACGGCATCCA CTTCTGCATC GCGCCCCCTT TGGCCCGGTT GGAGGCCCGG ATCGCCGTCC
2461 GCGCCCTTCT CGAACGCTGC CCGGACCTCG CCCTGGACGT CTCCCCCGGC GAACTCGTGT
2521 GGTATCCGAA CCCGATGATC CGCGGGCTCA AGGCCCTGCC GATCCGCTGG CGGCGAGGAC
5 2581 GGGAGGCGGG CCGCCGTACC GGTGAACCC GCACGTCACC CATTACGACT CCTTGTACAG
2641 GAAGCCCCGG ATCGGTCCCC CCTCGCCGTA ACAAGACCTG GTTAGAGTGA TGGAGGACGA
2701 CGAAGGGTTC GGCGCCCGGA CGAGGGGGGA CTTCCGCGAT GAATCTGGTG GAACGCGACG
2761 GGGAGATAGC CCATCTCAGG GCCGTTCTTG ACGCATCCGC CGCAGGTGAC GGGACGCTCT
10 2821 TACTCGTCTC CGGACCGGCC GGCAGCGGGA AGACGGAGCT GCTGCGGTCT CTCCGCCGGC
2881 TGGCCGCCGA GCGGGAGACC CCCGTCTGGT CGGTCCGGGC GCTGCCGGGT GACCGCGACA
2941 TCCCCCTGGG CGTCTCTGTC CAGTTACTCC GCAGCGCCGA ACAACACGGT GCCGACACCT
3001 CCGCCGTCCG CGACCTGCTG GACGCCGCCT CGCGGCGGGC CGGAACCTCA CCTCCCCCGC
3061 CGACGCGCCG CTCCGCGTCG ACGAGACACA CCGCCTGCAC GACTGGCTGC TCTCCGTCTC
3121 CCGCCGGCAC CCCGTTCTCT GTCGCCGTCG ACGACCTGAC CCACGCCGAC ACCGCGTCCC
15 3181 TGAGGTTCTT CCTGTACTGC GCCGCCACC ACGACCAGGG CGGCATCGGC TTCGTATGA
3241 CCGAGCGGGC CTCGAGCGC GCCGGATACC GGGTGTTCCG CGCCGAGCTG CTCCGCCAGC
3301 CGCACTGCCG CAACATGTGG CTCTCCGGGC TTCCCCCAG CGGGGTACGC CAGTTACTCG
3361 CCCACTACTA CGGCCCCGAG GCCGCCGAGC GCGGGGCCCC CGCGTACCAC GCGACGACCG
3421 GCGGGAACCC GCTGCTCTCT CGGGCGCTGA CCCAGGACCG GCAGGCCTCC CACACCACCC
20 3481 TCGGCGCGGC CCGCGGCGAC GAGCCCGTCC ACGGCGACGC CTTGCCCCAG GCCGTCTCTG
3541 ACTGCCTGCA CCGCAGCGCC GAGGGCACAC TGGAGACCGC CCGCTGGCTC GCGGTCTCTG
3601 AACAGTCCGA CCCGCTCTCT GTGGAGCGGC TCACGGGAAC GACCGCCGCC GCCGTGAGC
3661 GCCACATCCA GGAGCTCGCC GCCATCGGCC TCCTGGACGA GGACGGCACC CTGGGACAGC
3721 CCGCGATCCG CGAGGCGGCC CTCCAGGACC TGCCGGCCGG CGAGCGCACC GAACTGCACC
25 3781 GCGCGCGCCG GGAGCAGCTG CACCGGGACG GCGCCGACGA GGACACCGTG GCCCGCCACC
3841 TGCTGGTCGG CCGCGCCCCC GACGCTCCCT GGGCGCTGCC CCTGCTCGAA CGGGGCGCGC
3901 AGCAGGCCCT GTTCGACGAC CACTCGACG ACGCCTTCCG GATCCTCGAG TTCGCCGTGC
3961 GGTCGAGCAC CGACAACACC CAGCTGGCCC GCCTCGCCCC ACACCTGGTC GCGGCCTCCT
4021 GGCGGATGAA CCCGCACATG ACGACCCGGG CCCTCGCACT CTTGACCCGG CTCCTGAGCG
30 4081 GTGAACTGCC GCCCAGCCAC CCGGTCATGG CCCTGATCCG CTGCCTCGTC TGGTACGGNC
4141 GGCTGCCCGA GGCCGCCGAC GCGCTGTCCC GGCTGCGGCC CAGCTCCGAC AACGATGCCT
4201 TGGAGCTGTC GCTCACCCTG ATGTGGCTCG CGGCGCTGTG CCCGCCGCTC CTGGAGTCCC
4261 TGCCGGCCAC GCCGGAGCCG GAGCGGGGTC CCGTCCCCGT ACGGCTCGCG CCGCGGACGA
4321 CCGCGCTCCA GGCCAGGCC GCGCTCTTCC AGCGGGGCCC GGACAACGCC TCGGTGCGCG
35 4381 AGGCCGAACA GATCCTGCAG GGCTGCCGGC TGTCGGAGGA GACGTACGAG GCCCTGGAGA
4441 CCGCCCTCTT GGTCTCTGTC CACGCCGACC GGCTCGACCG GCGCTGTTC TGGTGGACG
4501 CCCTGCTCGC CGAGGCCGTG GAGCGGCGGT CGCTCGGCTG GGAGGCGGTC TTCGCCGCGA
4561 CCCGGGCGAT GATCGCGATC CGTGCGGCG ACCTCCCGAC GCGCGGGGAG CGGGCCGAGC
4621 TGGCGCTCTC CCACGCGGCG CCGGAGAGCT GGGGCCCTCG CGTGGGCATG CCCCTCTCCG
40 4681 CGCTGCTGCT CGCCTGCACG GAGGCCGGCG AGTACGAACA GCGGAGCGG GTCCTGCGGC
4741 AGCCGGTGCC GGACGCGATG TTCGACTCGC GGCACGGCAT GGAGTACATG CACGCCCGGG
4801 GCCGCTACTG GCTGGCGANC GGCCGGCTGC ACGCGGCGCT GGGCGAGTTC ATGCTCTGCG
4861 GGGAGATCCT GGGCAGCTGG AACCTCGACC AGCCCTCGAT CGTGCCCTGG CGGACCTCCG
4921 CCGCCGAGGT GTACCTGCGG CTCGGCAACC GCCAGAAGGC CAGGGCGCTG GCCGAGGCC
45 4981 AGCTCGCCCT GGTGCGGCCC GGGCGCTCCC GCACCCGGGG TCTACCCTG CGGTCTCTGG
5041 CCGCGGCGGT GGACGGCCAG CAGGCGGAGC GGCTGCACGC CGAGGCGGTC GACATGCTGC
5101 ACGACAGCGG CGACCGGCTC GAACACGCCC GCGCGCTCGC CGGGATGAGC CGCCACCAGC
5161 AGGCCAGGG GGACAACCTAC CGGGCGAGGA TGACGGCGCG GCTCGCCGGC GACATGGCGT
5221 GGGCCTGCGG CGCGTACCCG CTGGCCGAGG AGATCGTGCC GGGCCGCGGC GGCCGCCGGG
50 5281 CGAAGGCGGT GAGCACGGAG CTGGAACCTG CCGGCGGCCC GGACGTGCGC CTGCTCTCGG
5341 AGGCCGAACG CCGGGTGGCG GCCCTGGCAG CCCGAGGATT GACGAACCGC CAGATAGCGC
5401 GCCGGCTCTG CGTCACCGCG AGCACGGTCG AACAGCACCT GACGCGCGTC TACCGCAAAC
5461 TGAACGTGAC CCGCCGAGCA GACCTCCCGA TCAGCCTCGC CCAGGACAAG TCCGTACCGG
5521 CCTGAGCCAC CCCCCTGTCT CCCGTGCGAC GACCCGCGGC ACGGGCCACC GGGCCCGCCG
55 5581 GGACACGCGG GTGCGACACG GGGGCGCGCC AGGTGCCATG GGGACCTCCG TGACCGCCCC
5641 AGGCGCCCGA GGCGCCCGGT GCGGCACCCG GAGACGCCAG GACCGCCGGG ACCACCGGAG
5701 ACGCCAGGGA CCGCTGGGGA CACCGGGACC TCAGGGACCG CCGGGACCGC CCGAGTTGCA
5761 CCCGGTGCGC CCGGGGACAC CAGACCGCCG GGACCAACCG AGGGTGCCCG GTGTGGCCCC
5821 GGCGGCCCGG GTGTCCTTCA TCGGTGGGCC TTCATCGGCA GGAGGAAGCG ACCGTGAGAC

5881 CCGTCGTGCC GTCGGCGATC AGCCGCCTGT ACGGGCGTCG GACTCCCTGG CCGTCCCGGA
5941 CCCGTCGTAC GGGCTCGCGG GACCCGGTGC (SEQ ID NO:21)

Contig 003 from cosmid pKOS023-26 contains 3292 nucleotides and the following

- 5 ORFs: from nucleotide 104 to 982 is ORF13, which encodes dNDP glucose synthase (glucose-1-phosphate thymidyl transferase); from nucleotide 1114 to 2127 is ORF14, which encodes dNDP-glucose 4,6-dehydratase; and from nucleotide 2124 to 3263 is the *picC1* ORF. (SEQ ID NO:22)

10 1 ACCCCCCAAA GGGGTGGTGA CACTCCCCCT GCGCAGCCCC TAGCGCCCCC CTAAGTCGCC
61 ACGCCGACCG TTATCACCAG CGCCCTGCTG CTAGTTTCCG AGAATGAAGG GAATAGTCCT
121 GGCCGGCGGG AGCGGAACTC GGCTGCATCC GCGGACCTCG GTCATTTCGA AGCAGATTCT
181 TCCGGTCTAC AACAAACCGA TGATCTACTA TCCGCTGTCG GTTCTCATGC TCGGCGGTAT
241 TCGCGAGATT CAAATCATCT CGACCCCCCA GCACATCGAA CTCTTCCAGT CGCTTCTCGG
15 301 AAACGGCAGG CACCTGGGAA TAGAACTCGA CTATGCGGTC CAGAAAGAGC CCGCAGGAAT
361 CGCGGACGCA CTTCTCGTCG GAGCCGAGCA CATCGGCGAC GACACCTGCG CCCTGATCCT
421 GGGCGACAAC ATCTTCCACG GGCCCGGCCT CTACACGCTC CTGCGGGACA GCATCGCGCG
481 CCTCGACGGC TCGGTGCTCT TCGGCTACCC GGTCAAGGAC CCCGAGCGGT ACGGCGTCGC
541 CGAGGTGGAC GCGACGGGCC GGCTGACCGA CCTCGTCGAG AAGCCCGTCA AGCCGCGCTC
601 CAACCTCGCC GTCACCGGCC TCTACCTCTA CGACAACGAC GTCGTCGACA TCGCCAAGAA
20 661 CATCCGGCCC TCGCCGCGCG GCGAGCTGGA GATCACCAGC GTCAACCGCG TCTACCTGGA
721 GCGGGGCGCG GCCGAACCTG TCAACCTGGG CCGCGGCTTC GCCTGGCTGG ACACCGGCAC
781 CCACGACTCG CTCCTGCGGG CCGCCAGTA CGTCCAGGTC CTGGAGGAGC GGCAGGGCGT
841 CTGGATCGCG GGCCTTGAGG AGATCGCCTT CCGCATGGGC TTCATCGACG CCGAGGCCTG
901 TCACGGCCTG GGAGAAGGCC TCTCCCGCAC CGAGTACGGC AGCTATCTGA TGGAGATCGC
25 961 CGGCCGCGAG GGAGCCCCGT GAGGGCACCT CGCGGCGGAC GCGTTCCAC GACCGACAGC
1021 GCCACCGACA GTGCGACCCA CACCGCGACC CGCACCGCCA CCGACAGTGC GACCCACACC
1081 GCGACCTACA GCGCGACCGA AAGGAAGACG GCAGTGCAGG TTCTGGTGAC CGGAGGTGCG
1141 GGCTTCATCG GCTCGCACTT CGTGCGGCGC CTCCTCGCCG GGGCGTACCC CGACGTGCCC
1201 GCGGATGAGG TGATCGTCCT GGACAGCCTC ACCTACGCGG GCAACCGCGC CAACCTCGCC
30 1261 CCGGTGGACG CGGACCGCG ACTGCGCTTC GTCCACGGCG ACATCCGCGA CGCCGGCCTC
1321 CTCGCCCCGG AACTGCGCGG CGTGAGCGCC ATCGTCCACT TCGCGGCCGA GAGCCACGTG
1381 GACCGCTCCA TCGCGGGCGC GTCCGTGTTT ACCGAGACCA ACGTGCAGGG CACGCAGACG
1441 CTGCTCCAGT GCGCCGTCGA CGCCGGCGTC GGCCGGGTCG TGCACGTCTC CACCGACGAG
1501 GTGTACGGGT CGATCGACTC CGGCTCCTGG ACCGAGAGCA GCCCCTGGA GCCCAACTCG
35 1561 CCCTACGCGG CGTCCAAGGC CGGCTCCGAC CTCGTTGCCC GCGCCTACCA CCGGACGTAC
1621 GGCTTCGACG TACGGATCAC CCGCTGCTGC AACAACTACG GGCCGTACCA GCACCCCGAG
1681 AAGCTCATCC CCCTCTTCGT GACGAACCTC CTCGACGGCG GGACGCTCCC GCTGTACGGC
1741 GACGGCGCGA ACGTCCGCGA GTGGGTGCAC ACCGACGACC ACTGCCGGGG CATCGCGCTC
1801 GTCTTCGCGG GCGGCCGGGC CGGCGAGATC TACCACATCG GCGGCGGCCT GGAGCTGACC
40 1861 AACCAGCAAC TCACCGGCAT CCTCCTGGAC TCGCTCGGCG CCGACTGGTC CTCGGTCCGG
1921 AAGGTCGCCG ACCGCAAGGG CCACGACCTG CGTACTCCC TCGACGGCGG CAAGATCGAG
1981 CGCGAGCTCG GCTACCGCCC GCAGGTCTCC TTCGCGGACG GCCTCGCGCG GACCGTCCGC
2041 TGGTACCGGG AGAACCAGCG CTGGTGGGAG CCGCTCAAGG CGACCGCCCC GCAGTGGCC
2101 GCCACCGCCG TGGAGGTGTC CGCGTGAGCA GCGCGGCCGA GACCCCGCGC GTCCCTTCC
45 2161 TCGACCTCAA GGCCGCTAC GAGGAGCTCC GCGCGGAGAC CGACGCGCG ATCGCCCGCG
2221 TCCTCGACTC GGGGCGCTAC CTCCTCGGAC CCGAACTCGA AGGATTCGAG GCGGAGTTCG
2281 CCGCGTACTG CGAGACGGAC CACGCCGTCG GCGTGAACAG CCGGATGGAC GCCCTCCAGC
2341 TCGCCCTCCG CGGCCTCGGC ATCGGACCCG GGGACGAGGT GATCGTCCCC TCGCACACGT
2401 ACATCGCCAG CTGGCTCGCG GTGTCCGCCA CCGGCGCGAC CCGCGTGCCC GTCGAGCCGC
50 2461 ACGAGGACCA CCCACCCCTG GACCCGCTGC TCGTCGAGAA GGCGATCACC CCCCGCACCC
2521 GGGCGCTCCT CCCCGTCCAC CTCTACGGGC ACCCGCCGA CATGGACGCC CTCCGCGAGC
2581 TCGCGGACCG GCACGGCCTG CACATCGTCG AGGACGCGCG GCAGGCCAC GCGCGCCGCT
2641 ACCGGGGCCG GCGGATCGGC GCCGGGTCGT CGGTGGCCCG GTTCAGCTTC TACCCGGGCA
2701 AGAACCTCGG CTGCTTCGGC GACGGCGGCG CCGTCGTCAC CCGCGACCCC GAGCTCGCCG

2761 AACGGCTCCG GATGCTCCGC AACTACGGCT CGCGGCAGAA GTACAGCCAC GAGACGAAGG
 2821 GCACCAACTC CCGCCTGGAC GAGATGCAGG CCGCCGTGCT GCGGATCCGG CTCGNCCACC
 2881 TGGACAGCTG GAACGGCCGC AGGTCGGCGC TGGCCGCGGA GTACCTCTCC GGGCTCGCCG
 2941 GACTGCCCCG CATCGGCCTG CCGGTGACCG CGCCCGACAC CGACCCGGTC TGGCACCTCT
 5 3001 TCACCGTGCG CACCGAGCGC CGCGACGAGC TGCGCAGCCA CCTCGACGCC CGCGGCATCG
 3061 ACACCCTCAC GCACTACCCG GTACCCGTGC ACCTCTCGCC CGCCTACGCG GCGGAGGCAC
 3121 CGCCGGAAGG CTCGCTCCCG CGGGCCGAGA GCTTCGCGCG GCAGGTCTCT AGCCTGCCGA
 3181 TCGGCCCGCA CCTGGAGCGC CCGCAGGCGC TGCGGGTGAT CGACGCCGTG CGCGAATGGG
 3241 CCGAGCGGGT CGACCAGGCC TAGTCAGGTG GTCCGGTAGA CCCAGCAGGC CG (SEQ ID
 10 NO:22)

Contig 004 from cosmid pKOS023-26 contains 1693 nucleotides and the following
 ORFs: from nucleotide 1692 to 694 is ORF15, which encodes a part of S-adenosylmethionine
 synthetase; and from nucleotide 692 to 1 is ORF16, which encodes a part of a protein

15 homologous to the *M. tuberculosis cbhK* gene. (SEQ ID NO:23)

1 ATGCGGCACC CCTTGGCGCC GAGCGTGGTG ATCCAGGTGC CGACCCGGGC GAGCACCTCC
 61 TGCTCGGTCC AGCCCGTCTT GCTGAGCAGC AGCGCCCGCT CGTAGGCGTT CGTGAACAGC
 121 AGCTCGGCTC CGTCGACGAG CTCCCGGACG CTGTCGCCCT CCAGCCGGGC GAGCTGCTGC
 181 GAGGGGTCCG CGGCCCCGGC GAGGCCCAGC TCGCGGCAGA CCCGCGTGTG CCGCACCATC
 20 241 GCCTCGGGGT CGTCCGCGCC GACGAGGACG AGGTCGATCC CGCCGGGCCG GCCGGCCGTC
 301 TCGCCAGGT CGATGTCGCG CGCCTCGGCC ATCGCGCCCG CGTAGAACGA GCGGAGCTGA
 361 TTGCCGTCCT CGTCGGTGGT GCACATGAAG CGGGCGGTGT GCTGACGGTC CGACACCCGC
 421 ACGGAGTCGG TGTCGACGCC CGCGGCGCGG AGCAGCTGCC CGTACCCGTC GAAGTCCTTG
 481 CCGACGGCGC CGACGAGGAC GGGGCGGCGA CCGAGCAGGC CGAGGCCGTA CGCGATGTTG
 25 541 GCGGCGACGC CGCCGTGCCG GATGTCCAGG GTGTCGACGA GGAACGACAG GGACACGTGG
 601 GCGAGCTGGT CCGGCAGGAT CTGCTCGGCG AAGCGGCCCG GGAAGGTCAT CAGGTGGTCG
 661 GTGGCGATCG ACCCGGTGAC GGCTATACGC ATGTCAGAGC CCCGCGGCCT TCTTCAGGGC
 721 GTCCACGCGG TCGGTGCGCT CCCAGGTGAA GTCCGGCAGC TCGCGGCCGA AGTGGCCGTA
 781 GCGGCGGTC TGGGAGTAGA TCGGGCGGAG CAGGTCGAGG TCGCGGATGA TCGCGGCCG
 30 841 GCGGAGGTCG AAGACCTCGC CGATGGCGTT CTCGATCTTC TCGGTCTCGA TCTTGTGGGT
 901 GCCGAAGGTC TCGACGAAGA GGCCGACGGG CTCGGCCTTG CCGATCGCGT ACGCGACCTG
 961 GACCTCGCAG CGCGAGGCGA GACCGGCGGC GACGACGTTT TTCGCCACCC AGCGCATCGC
 1021 GTACGCGGCG GAGCGGTCGA CCTTCGACGG GTCCTTGCCG GAGAAGGCGC CGCCACCGTG
 1081 GCGGGCCATG CCGCCGTAGG TGTCGATGAT GATCTTGCGG CCGGTGAGGC CGGCGTCGCC
 35 1141 CATCGGGCCG CCGATCTCGA AGCGACCGGT CGGGTTCACG AGCAGGCGGT AGCCGTCCGT
 1201 GTCGAGCTTG ATGCCGTCCT CGACGAGCTG CGCAAGCACG TGCTCGACGA CGAACTCCG
 1261 CACGTCGGGG GCGAGCAGCG ACTCCAGGTC GATGTCCGAG GCGTGCTGCG AGGAGACGAC
 1321 GACCGTGTCG AGACGGACCG CCCTGTCGCC GTCGTA CTG ATGGTGACCT GGGTCTTGCC
 1381 GTCGGGACGC AGGTACGGGA TGGTCCCGTT CTTGCGGACC TCGGTGAGGC GGCGCGAGAG
 40 1441 ACGGTGCGCG AGGTGGATCG GCAGCGGCAT CAGCTCGGGC GTCTCGTCCG AGGCATAGCC
 1501 GAACATCAGG CCCTGGTCAC CGGCGCCCTG CTTGTGAGC TCGTCCCCCT CGTCCCGCTG
 1561 GGAGGCACCC TCGACCCGCT TCTCGTACGC GGTGTGACA CCCTGGGCGA TGTCCGGGGA
 1621 CTGCGACCCG ATGGACACCG ACACGCCGCA GGAGGCGCCG TCGAAGCCCT TCTTCGAGGA
 1681 GTCGTACCCG ATC (SEQ ID NO:23)
 45

Contig 005 from cosmid pKOS023-26 contains 1565 nucleotides and contains the
 ORF of the *picCV* gene that encodes PICCV, involved in desosamine biosynthesis. (SEQ ID
 NO:24)

1 CCCCCTCGC GGCCCCCAG ACATCCACGC CCACGATTGG ACGTCCCGA TGACCGCCCC
 50 61 CGCCCTCTCC GCCACCGCCC CGGCCGAACG CTGCGCGCAC CCCGGAGCCG ATCTGGGGGC
 121 GGCGGTCCAC GCCGTCGGCC AGACCCTCGC CGCCGGCGGC CTCGTGCCGC CCGACGAGGC
 181 CGGAACGACC GCCCGCCACC TCGTCCGGCT CGCCGTGCGC TACGGCAACA GCCCCTTCAC

241 CCCGCTGGAG GAGGCCCGCC ACGACCTGGG CGTCGACCGG GACGCCTTCC GCGCCTCCT
 301 CGCCCTGTTC GGGCAGGTCC CGGAGCTCCG CACCGCGGTC GAGACCGGCC CCGCCGGGGC
 361 GTACTGGAAG AACACCCTGC TCCCGCTCGA ACAGCGCGGC GTCTTCGACG CGGCGCTCGC
 421 CAGGAAGCCC GTCTTCCCGT ACAGCGTCGG CCTCTACCCC GGCCCGACCT GCATGTTCCG
 5 481 CTGCCACTTC TGCCTCCGTG TGACCGGCGC CCGCTACGAC CCGTCCGCCC TCGACGCCGG
 541 CAACGCCATG TTCCGGTCGG TCATCGACGA GATACCCGCG GGCAACCCCT CGGCGATGTA
 601 CTTCTCCGGC GGCCTGGAGC CGCTCACCAG CCCCGGCCCT GGGAGCCTGG CCGCGCACGC
 661 CACCGACCAC GGCCTGCGGC CCACCGTCTA CACGAACTCC TTCGCGCTCA CCGAGCGCAC
 721 CCTGGAGCGC CAGCCCGGCC TCTGGGGCCT GCACGCCATC CGCACCTCGC TCTACGGCCT
 10 781 CAACGACGAG GAGTACGAGC AGACCACCGG CAAGAAGGCC GCCTTCCGCC GCGTCCGCGA
 841 GAACCTGCGC CGCTTCCAGC AGCTGCGCGC CGAGCGCGAG TCGCCGATCA ACCTCGGCTT
 901 CGCCTACATC GTGCTCCCGG GCGGTGCCTC CCGCCTGCTC GACCTGGTCG ACTTCATCGC
 961 CGACCTCAAC GACGCCGGGC AGGGCAGGAC GATCGACTTC GTCAACATTC GCGAGGACTA
 1021 CAGCGGCCGT GACGACGGCA AGCTGCCGCA GGAGGAGCGG GCCGAGCTCC AGGAGGCCCT
 15 1081 CAACGCCTTC GAGGAGCGGG TCCGCGAGCG CACCCCGGGA CTCCACATCG ACTACGGCTA
 1141 CGCCCTGAAC AGCCTGCGCA CCGGGGCCGA CGCCGAAGTG CTGCGGATCA AGCCCGCCAC
 1201 CATGCGGCCC ACCGCGCACC CGCAGGTCGC GGTGCAGGTC GATCTCCTCG GCGACGTGTA
 1261 CCTGTACCGC GAGGCCGGCT TCCCGACCT GGACGGCGCG ACCCGCTACA TCGCGGGCCG
 1321 CGTGACCCCC GACACCTCCC TCACCGAGGT CGTCAGGGAC TTCGTCGAGC GCGGCGGCGA
 20 1381 GGTGGCGGCC GTCGACGGCG ACGAGTACTT CATGGACGGC TTCGATCAGG TCGTCACCGC
 1441 CCGCCTGAAC CAGCTGGAGC GCGACGCCGC GGACGGCTGG GAGGAGGCCC GCGGCTTCCT
 1501 GCGCTGACCC GCACCCGCCC CGATCCCCC GATCCCCC CCACGATCCC CCCACCTGAG
 1561 GGCCC (SEQ ID NO:24)

25 The recombinant desosamine biosynthesis and transfer and beta-glucosidase genes and proteins provided by the invention are useful in the production of glycosylated polyketides in a variety of host cells, as described in Section IV below.

Section III. The Genes for Macrolide Ring Modification: the *picK* Hydroxylase Gene

30 The present invention provides the *picK* gene in recombinant form as well as recombinant PicK protein. The availability of the hydroxylase encoded by the *picK* gene in recombinant form is of significant benefit in that the enzyme can convert narbomycin into picromycin and accepts in addition a variety of polyketide substrates, particularly those related to narbomycin in structure. The present invention also provides methods of
 35 hydroxylating polyketides, which method comprises contacting the polyketide with the recombinant PicK enzyme under conditions such that hydroxylation occurs. This methodology is applicable to large numbers of polyketides.

DNA encoding the *picK* gene can be isolated from cosmid pKOS023-26 of the invention. The DNA sequence of the *picK* gene is shown in the preceding section. This DNA
 40 sequence encodes one of the recombinant forms of the enzyme provided by the invention. The amino acid sequence of this form of the *picK* gene is shown below. The present invention also provides a recombinant *picK* gene that encodes a *picK* gene product in which

the PicK protein is fused to a number of consecutive histidine residues, which facilitates purification from recombinant host cells.

Amino acid sequence of picromycin/methymycin cytochrome P450 hydroxylase, PicK (SEQ ID NO:18)

5
1 VVRTQQGTTA SPPVLDLGAL GQDFAADPYP TYARLRAEGP AHRVRTPEGD EVWLVVGYDR
61 ARAVLADPRF SKDWRNSTTP LTEAEAALNH NMLESPPRH TRLRKLVAE FTMRRVELLR
121 PRVQEIVDGL VDAMLAAPDG RADLMESLAW PLPITVISEL LGVPEPDRAA FRVWTDAFVF
181 PDDPAQAQTA MAEMSGYLSR LIDSKRGQDG EDLLSALVRT SDEDGSRLTS EELGMAHIL
10 241 LVAGHETTVN LIANGMYALL SHPDQLAALR ADMTLLDGAV EEMLRYEGPV ESATYRFPVE
301 PVDLDGTVIP AGDTVLVLA DAHRTPERFP DPHRFDIRRD TAGHLAFGHG IHFCIGAPLA
361 RLEARIAVRA LLERCPDLAL DVSPGELVWY PNPIMIRGLKA LPIRWRRGRE AGRRTG (SEQ ID NO:18)

15 The recombinant PicK enzyme of the invention hydroxylates narbomycin at the C12 position and YC-17 at either the C10 or C12 position. Hydroxylation of these compounds at the respective positions increases the antibiotic activity of the compound relative to the unhydroxylated compound. Hydroxylation can be achieved by a number of methods. First, the hydroxylation may be performed *in vitro* using purified hydroxylase, or the relevant
20 hydroxylase can be produced recombinantly and utilized directly in the cell that produces it. Thus, hydroxylation may be effected by supplying the nonhydroxylated precursor to a cell that expresses the hydroxylase. These and other details of this embodiment of the invention are described in additional detail below in Section IV and the examples.

25 Section IV: Heterologous Expression of the Narbonolide PKS; the Desosamine Biosynthetic and Transferase Genes; the Beta-Glucosidase Gene; and the *picK* Hydroxylase Gene

In one important embodiment, the invention provides methods for the heterologous expression of one or more of the genes involved in picromycin biosynthesis and recombinant DNA expression vectors useful in the method. Thus, included within the scope of the
30 invention in addition to isolated nucleic acids encoding domains, modules, or proteins of the narbonolide PKS, glycosylation, and/or hydroxylation enzymes, are recombinant expression systems. These systems contain the coding sequences operably linked to promoters, enhancers, and/or termination sequences that operate to effect expression of the coding sequence in compatible host cells. The host cells are modified by transformation with the
35 recombinant DNA expression vectors of the invention to contain these sequences either as extrachromosomal elements or integrated into the chromosome. The invention also provides

methods to produce PKS and post-PKS tailoring enzymes as well as polyketides and antibiotics using these modified host cells.

As used herein, the term expression vector refers to a nucleic acid that can be introduced into a host cell or cell-free transcription and translation medium. An expression
5 vector can be maintained stably or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm. An expression vector also comprises a gene that serves to produce RNA, which typically is translated into a polypeptide in the cell or cell extract. To drive production of the RNA, the expression vector typically comprises one or more promoter elements.
10 Furthermore, expression vectors typically contain additional functional elements, such as, for example, a resistance-conferring gene that acts as a selectable marker.

The various components of an expression vector can vary widely, depending on the intended use of the vector. In particular, the components depend on the host cell(s) in which the vector will be introduced or in which it is intended to function. Components for
15 expression and maintenance of vectors in *E. coli* are widely known and commercially available, as are components for other commonly used organisms, such as yeast cells and *Streptomyces* cells.

One important component is the promoter, which can be referred to as, or can be included within, a control sequence or control element, which drives expression of the
20 desired gene product in the heterologous host cell. Suitable promoters include those that function in eucaryotic or procaryotic host cells. In addition to a promoter, a control element can include, optionally, operator sequences, and other elements, such as ribosome binding sites, depending on the nature of the host. Regulatory sequences that allow for regulation of expression of the heterologous gene relative to the growth of the host cell may also be
25 included. Examples of such regulatory sequences known to those of skill in the art are those that cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus.

Preferred host cells for purposes of selecting vector components include fungal host cells such as yeast and procaryotic, especially *E. coli* and *Streptomyces*, host cells, but single
30 cell cultures of, for example, mammalian cells can also be used. In hosts such as yeasts, plants, or mammalian cells that ordinarily do not produce polyketides, it may be necessary to provide, also typically by recombinant means, suitable holo-ACP synthases to convert the recombinantly produced PKS to functionality. Provision of such enzymes is described, for

example, in PCT publication Nos. WO 97/13845 and WO 98/27203, each of which is incorporated herein by reference. Control systems for expression in yeast, including controls that effect secretion are widely available and can be routinely used. For *E. coli* or other bacterial host cells, promoters such as those derived from sugar metabolizing enzymes, such as galactose, lactose (lac), and maltose, can be used. Additional examples include promoters derived from genes encoding biosynthetic enzymes, and the tryptophan (trp), the beta-lactamase (bla), bacteriophage lambda PL, and T5 promoters. In addition, synthetic promoters, such as the tac promoter (U.S. Patent No. 4,551,433), can also be used.

Particularly preferred are control sequences compatible with *Streptomyces* spp. Particularly useful promoters for *Streptomyces* host cells include those from PKS gene clusters that result in the production of polyketides as secondary metabolites, including promoters from aromatic (Type II) PKS gene clusters. Examples of Type II PKS gene cluster promoters are act gene promoters and *tcm* gene promoters; an example of a Type I PKS gene cluster promoter is the spiramycin PKS gene promoter.

If a *Streptomyces* or other host ordinarily produces polyketides, it may be desirable to modify the host so as to prevent the production of endogenous polyketides prior to its use to express a recombinant PKS of the invention. Such hosts have been described, for example, in U.S. Patent No. 5,672,491, incorporated herein by reference. In such hosts, it may not be necessary to provide enzymatic activities for all of the desired post-translational modifications of the enzymes that make up the recombinantly produced PKS, because the host naturally expresses such enzymes. In particular, these hosts generally contain holo-ACP synthases that provide the pantotheinyl residue needed for functionality of the PKS.

Thus, in one important embodiment, the vectors of the invention are used to transform *Streptomyces* host cells to provide the recombinant *Streptomyces* host cells of the invention. *Streptomyces* is a convenient host for expressing narbonolide or 10-deoxymethynolide or derivatives of those compounds, because narbonolide and 10-deoxymethynolide are naturally produced in certain *Streptomyces* species, and *Streptomyces* generally produce the precursors needed to form the desired polyketide. The present invention also provides the narbonolide PKS gene promoter in recombinant form, located upstream of the *picAI* gene on cosmid pKOS023-27. This promoter can be used to drive expression of the narbonolide PKS or any other coding sequence of interest in host cells in which the promoter functions, particularly *S. venezuelae* and generally any *Streptomyces* species. As described below, however,

promoters other than the promoter of the narbonolide PKS genes will typically be used for heterologous expression.

For purposes of the invention, any host cell other than *Streptomyces venezuelae* is a heterologous host cell. Thus, *S. narbonensis*, which produces narbomycin but not picromycin
5 is a heterologous host cell of the invention, although other host cells are generally preferred for purposes of heterologous expression. Those of skill in the art will recognize that, if a *Streptomyces* host that produces a picromycin or methymycin precursor is used as the host cell, the recombinant vector need drive expression of only a portion of the genes constituting the picromycin gene cluster. As used herein, the picromycin gene cluster includes the
10 narbonolide PKS, the desosamine biosynthetic and transferase genes, the beta-glucosidase gene, and the *picK* hydroxylase gene. Thus, such a vector may comprise only a single ORF, with the desired remainder of the polypeptides encoded by the picromycin gene cluster provided by the genes on the host cell chromosomal DNA.

The present invention also provides compounds and recombinant DNA vectors useful
15 for disrupting any gene in the picromycin gene cluster (as described above and illustrated in the examples below). Thus, the invention provides a variety of modified host cells (particularly, *S. narbonensis* and *S. venezuelae*) in which one or more of the genes in the picromycin gene cluster have been disrupted. These cells are especially useful when it is desired to replace the disrupted function with a gene product expressed by a recombinant
20 DNA vector. Thus, the invention provides such *Streptomyces* host cells, which are preferred host cells for expressing narbonolide derivatives of the invention. Particularly preferred host cells of this type include those in which the coding sequence for the loading module has been disrupted, those in which one or more of any of the PKS gene ORFs has been disrupted, and/or those in which the *picK* gene has been disrupted.

25 In a preferred embodiment, the expression vectors of the invention are used to construct a heterologous recombinant *Streptomyces* host cell that expresses a recombinant PKS of the invention. As noted above, a heterologous host cell for purposes of the present invention is any host cell other than *S. venezuelae*, and in most cases other than *S. narbonensis* as well. Particularly preferred heterologous host cells are those which lack
30 endogenous functional PKS genes. Illustrative host cells of this type include the modified *Streptomyces coelicolor* CH999 and similarly modified *S. lividans* described in PCT publication No. WO 96/40968.

The invention provides a wide variety of expression vectors for use in *Streptomyces*. For replicating vectors, the origin of replication can be, for example and without limitation, a low copy number vector, such as SCP2* (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory manual* (The John Innes Foundation, Norwich, U.K., 1985);
5 Lydiate *et al.*, 1985, *Gene* 35: 223-235; and Kieser and Melton, 1988, *Gene* 65: 83-91, each of which is incorporated herein by reference), SLP1.2 (Thompson *et al.*, 1982, *Gene* 20: 51-62, incorporated herein by reference), and pSG5(ts) (Muth *et al.*, 1989, *Mol. Gen. Genet.* 219: 341-348, and Bierman *et al.*, 1992, *Gene* 116: 43-49, each of which is incorporated herein by reference), or a high copy number vector, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, *J.*
10 *Gen. Microbiol.* 129: 2703-2714; Vara *et al.*, 1989, *J. Bacteriol.* 171: 5782-5781; and Servin-Gonzalez, 1993, *Plasmid* 30: 131-140, each of which is incorporated herein by reference). High copy number vectors are generally, however, not preferred for expression of large genes or multiple genes. For non-replicating and integrating vectors and generally for any vector, it is useful to include at least an *E. coli* origin of replication, such as from pUC, p1P, p1I, and
15 pBR. For phage based vectors, the phage phiC31 and its derivative KC515 can be employed (see Hopwood *et al.*, *supra*). Also, plasmid pSET152, plasmid pSAM, plasmids pSE101 and pSE211, all of which integrate site-specifically in the chromosomal DNA of *S. lividans*, can be employed.

Preferred *Streptomyces* host cell/vector combinations of the invention include
20 *S. coelicolor* CH999 and *S. lividans* K4-114 host cells, which do not produce actinorhodin, and expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference.

As described above, particularly useful control sequences are those that alone or
25 together with suitable regulatory systems activate expression during transition from growth to stationary phase in the vegetative mycelium. The system contained in the illustrative plasmid pRM5, i.e., the *actI/actIII* promoter pair and the *actII-ORF4* activator gene, is particularly preferred. Other useful *Streptomyces* promoters include without limitation those from the *ermE* gene and the *melC1* gene, which act constitutively, and the *tipA* gene and the *merA*
30 gene, which can be induced at any growth stage. In addition, the T7 RNA polymerase system has been transferred to *Streptomyces* and can be employed in the vectors and host cells of the invention. In this system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of the chromosome or in a vector under the control of the inducible *merA*

promoter, and the gene of interest is placed under the control of the T7 promoter. As noted above, one or more activator genes can also be employed to enhance the activity of a promoter. Activator genes in addition to the *actII-ORF4* gene described above include *dnrI*, *redD*, and *ptpA* genes (see U.S. patent application Serial No. 09/181,833, *supra*).

- 5 Typically, the expression vector will comprise one or more marker genes by which host cells containing the vector can be identified and/or selected. Selectable markers are often preferred for recombinant expression vectors. A variety of markers are known that are useful in selecting for transformed cell lines and generally comprise a gene that confers a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium.
- 10 Such markers include, for example, genes that confer antibiotic resistance or sensitivity to the plasmid. Alternatively, several polyketides are naturally colored, and this characteristic can provide a built-in marker for identifying cells. Preferred selectable markers include antibiotic resistance conferring genes. Preferred for use in *Streptomyces* host cells are the *ermE* (confers resistance to erythromycin and lincomycin), *tsr* (confers resistance to thiostrepton), *aadA*
- 15 (confers resistance to spectinomycin and streptomycin), *aacC4* (confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), *hyg* (confers resistance to hygromycin), and *vph* (confers resistance to viomycin) resistance conferring genes.

To provide a preferred host cell and vector for purposes of the invention, the narbonolide PKS genes were placed on a recombinant expression vector that was transferred

20 to the non-macrolide producing host *Streptomyces lividans* K4-114, as described in Example 3. Transformation of *S. lividans* K4-114 with this expression vector resulted in a strain which produced two compounds in similar yield (~5-10 mg/L each). Analysis of extracts by LC/MS followed by ¹H-NMR spectroscopy of the purified compounds established their identity as narbonolide (Figure 5, compound 4) and 10-deoxymethynolide (Figure 5, compound 5), the

25 respective 14 and 12-membered polyketide precursors of narbomycin and YC17.

To provide a host cell of the invention that produces the narbonolide PKS as well as an additional narbonolide biosynthetic gene and to investigate the possible role of the PIC TEII in picromycin biosynthesis, the *picB* gene was integrated into the chromosome to provide the host cell of the invention *Streptomyces lividans* K39-18. The *picB* gene was

30 cloned into the *Streptomyces* genome integrating vector pSET152 (see Bierman *et al.*, 1992, *Gene* 116: 43, incorporated herein by reference) under control of the same promoter (*PactI*) as the PKS on plasmid pKOS039-86.

A comparison of strains *Streptomyces lividans* K39-18/pKOS039-86 and K4-114/pKOS039-86 grown under identical conditions indicated that the strain containing TEII produced 4-7 times more total polyketide. This increased production indicates that the enzyme is functional in this strain and is consistent with the observation that yields fall to
5 below 5% for both picromycin and methymycin when *picB* is disrupted in *S. venezuelae*. Because the production levels of compound 4 and 5 from K39-18/pKOS03986 increased by the same relative amounts, TEII does not appear to influence the ratio of 12 and 14-membered lactone ring formation. Thus, the invention provides methods of coexpressing the *picB* gene product or any other type II thioesterase with the narbonolide PKS or any other
10 PKS in heterologous host cells to increase polyketide production. However, transformation of a 6-dEB-producing *Streptomyces lividans*/pCK7 strain with an expression vector of the invention that produces PIC TEII resulted in little or no increase in 6-dEB levels, indicating that TEII enzymes may have some specificity for their cognate PKS complexes and that use of homologous TEII enzymes will provide
15 optimal activity.

In accordance with the methods of the invention, picromycin biosynthetic genes in addition to the genes encoding the PKS and PIC TEII can be introduced into heterologous host cells. In particular, the *picK* gene, desosamine biosynthetic genes, and the desosaminyl transferase gene can be expressed in the recombinant host cells of the invention to produce
20 any and all of the polyketides in the picromycin biosynthetic pathway (or derivatives thereof). Those of skill will recognize that the present invention enables one to select whether only the 12-membered polyketides, or only the 14-membered polyketides, or both 12- and 14-membered polyketides will be produced. To produce only the 12-membered polyketides, the invention provides expression vectors in which the last module is deleted or the KS domain
25 of that module is deleted or rendered inactive. If module 6 is deleted, then one preferably deletes only the non-TE domain portion of that module or one inserts a heterologous TE domain, as the TE domain facilitates cleavage of the polyketide from the PKS and cyclization and thus generally increases yields of the desired polyketide. To produce only the 14-membered polyketides, the invention provides expression vectors in which the
30 coding sequences of extender modules 5 and 6 are fused to provide only a single polypeptide.

In one important embodiment, the invention provides methods for desosaminylating polyketides or other compounds. In this method, a host cell other than *Streptomyces*

venezuelae is transformed with one or more recombinant vectors of the invention comprising the desosamine biosynthetic and desosaminyl transferase genes and control sequences positioned to express those genes. The host cells so transformed can either produce the polyketide to be desosaminylated naturally or can be transformed with expression vectors
5 encoding the PKS that produces the desired polyketide. Alternatively, the polyketide can be supplied to the host cell containing those genes. Upon production of the polyketide and expression of the desosamine biosynthetic and desosaminyl transferase genes, the desired desosaminylated polyketide is produced. This method is especially useful in the production of polyketides to be used as antibiotics, because the presence of the desosamine residue is
10 known to increase, relative to their undesosaminylated counterparts, the antibiotic activity of many polyketides significantly. The present invention also provides a method for desosaminylating a polyketide by transforming an *S. venezuelae* or *S. narbonensis* host cell with a recombinant vector that encodes a PKS that produces the polyketide and culturing the transformed cell under conditions such that said polyketide is produced and desosaminylated.
15 In this method, use of an *S. venezuelae* or *S. narbonensis* host cell of the invention that does not produce a functional endogenous narbonolide PKS is preferred.

In a related aspect, the invention provides a method for improving the yield of a desired desosaminylated polyketide in a host cell, which method comprises transforming the host cell with a beta-glucosidase gene. This method is not limited to host cells that have been
20 transformed with expression vectors of the invention encoding the desosamine biosynthetic and desosaminyl transferase genes of the invention but instead can be applied to any host cell that desosaminylates polyketides or other compounds. Moreover, while the beta-glucosidase gene from *Streptomyces venezuelae* provided by the invention is preferred for use in the method, any beta-glucosidase gene may be employed. In another embodiment, the beta-
25 glucosidase treatment is conducted in a cell free extract.

Thus, the invention provides methods not only for producing narbonolide and 10-deoxymethynolide in heterologous host cells but also for producing narbomycin and YC-17 in heterologous host cells. In addition, the invention provides methods for expressing the *picK* gene product in heterologous host cells, thus providing a means to produce picromycin,
30 methymycin, and neomethymycin in heterologous host cells. Moreover, because the recombinant expression vectors provided by the invention enable the artisan to provide for desosamine biosynthesis and transfer and/or C10 or C12 hydroxylation in any host cell, the invention provides methods and reagents for producing a very wide variety of glycosylated

and/or hydroxylated polyketides. This variety of polyketides provided by the invention can be better appreciated upon consideration of the following section relating to the production of polyketides from heterologous or hybrid PKS enzymes provided by the invention.

5 Section V: Hybrid PKS Genes

 The present invention provides recombinant DNA compounds encoding each of the domains of each of the modules of the narbonolide PKS, the proteins involved in desosamine biosynthesis and transfer to narbonolide, and the PicK protein. The availability of these compounds permits their use in recombinant procedures for production of desired portions of
10 the narbonolide PKS fused to or expressed in conjunction with all or a portion of a heterologous PKS. The resulting hybrid PKS can then be expressed in a host cell, optionally with the desosamine biosynthesis and transfer genes and/or the *picK* hydroxylase gene to produce a desired polyketide.

 Thus, in accordance with the methods of the invention, a portion of the narbonolide
15 PKS coding sequence that encodes a particular activity can be isolated and manipulated, for example, to replace the corresponding region in a different modular PKS. In addition, coding sequences for individual modules of the PKS can be ligated into suitable expression systems and used to produce the portion of the protein encoded. The resulting protein can be isolated and purified or can may be employed *in situ* to effect polyketide synthesis. Depending on the
20 host for the recombinant production of the domain, module, protein, or combination of proteins, suitable control sequences such as promoters, termination sequences, enhancers, and the like are ligated to the nucleotide sequence encoding the desired protein in the construction of the expression vector.

 In one important embodiment, the invention thus provides a hybrid PKS and the
25 corresponding recombinant DNA compounds that encode those hybrid PKS enzymes. For purposes of the invention, a hybrid PKS is a recombinant PKS that comprises all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a first PKS and all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a second PKS. In one preferred embodiment, the first PKS is
30 most but not all of the narbonolide PKS, and the second PKS is only a portion or all of a non-narbonolide PKS. An illustrative example of such a hybrid PKS includes a narbonolide PKS in which the natural loading module has been replaced with a loading module of another

PKS. Another example of such a hybrid PKS is a narbonolide PKS in which the AT domain of extender module 3 is replaced with an AT domain that binds only malonyl CoA.

In another preferred embodiment, the first PKS is most but not all of a non-narbonolide PKS, and the second PKS is only a portion or all of the narbonolide PKS. An
5 illustrative example of such a hybrid PKS includes a DEBS PKS in which an AT specific for methylmalonyl CoA is replaced with the AT from the narbonolide PKS specific for malonyl CoA.

Those of skill in the art will recognize that all or part of either the first or second PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For
10 example, only a small portion of an AT domain determines its specificity. See U.S. provisional patent application Serial No. 60/091,526, and Lau *et al.*, *infra*, incorporated herein by reference. The state of the art in DNA synthesis allows the artisan to construct *de novo* DNA compounds of size sufficient to construct a useful portion of a PKS module or domain. Thus, the desired derivative coding sequences can be synthesized using standard
15 solid phase synthesis methods such as those described by Jaye *et al.*, 1984, *J. Biol. Chem.* 259: 6331, and instruments for automated synthesis are available commercially from, for example, Applied Biosystems, Inc. For purposes of the invention, such synthetic DNA compounds are deemed to be a portion of a PKS.

With this general background regarding hybrid PKSs of the invention, one can better
20 appreciate the benefit provided by the DNA compounds of the invention that encode the individual domains, modules, and proteins that comprise the narbonolide PKS. As described above, the narbonolide PKS is comprised of a loading module, six extender modules composed of a KS, AT, ACP, and optional KR, DH, and ER domains, and a thioesterase domain. The DNA compounds of the invention that encode these domains individually or in
25 combination are useful in the construction of the hybrid PKS encoding DNA compounds of the invention.

The recombinant DNA compounds of the invention that encode the loading module of the narbonolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that
30 encodes the narbonolide PKS loading module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for the loading module of the heterologous PKS is replaced by that for the coding sequence of the narbonolide PKS loading module provides a novel PKS. Examples

include the 6-deoxyerythronolide B, rapamycin, FK506, FK520, rifamycin, and avermectin PKS coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the narbonolide PKS loading module is inserted into a DNA compound that comprises the coding sequence for the narbonolide PKS or a recombinant narbonolide PKS
5 that produces a narbonolide derivative in a different location in the modular system.

In another embodiment, a portion of the loading module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, replacing the propionyl CoA specific AT with an acetyl CoA, butyryl CoA, or other CoA specific AT. In addition, the KS^Q and/or ACP can be replaced by another
10 inactivated KS and/or another ACP. Alternatively, the KS^Q, AT, and ACP of the loading module can be replaced by an AT and ACP of a loading module such as that of DEBS. The resulting heterologous loading module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes narbonolide, a narbonolide derivative, or another polyketide.

15 The recombinant DNA compounds of the invention that encode the first extender module of the narbonolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the narbonolide PKS first extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting
20 construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the first extender module of the narbonolide PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the first extender module of the narbonolide PKS is inserted into a DNA compound
25 that comprises coding sequences for the narbonolide PKS or a recombinant narbonolide PKS that produces a narbonolide derivative or into a different location in the modular system.

In another embodiment, a portion or all of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific
30 AT with a malonyl CoA, ethylmalonyl CoA, or carboxyglycolyl CoA specific AT; deleting (which includes inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the

heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the narbonolide PKS, from a gene for a PKS that produces a polyketide other than narbonolide, or from chemical synthesis. The resulting heterologous first extender module coding sequence can be utilized in conjunction with a coding sequence
5 for a PKS that synthesizes narbonolide, a narbonolide derivative, or another polyketide.

In an illustrative embodiment of this aspect of the invention, the invention provides recombinant PKSs and recombinant DNA compounds and vectors that encode such PKSs in which the KS domain of the first extender module has been inactivated. Such constructs are especially useful when placed in translational reading frame with the remaining modules and
10 domains of a narbonolide PKS or narbonolide derivative PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the PKS encoded thereby can be fed or supplied with N-acetylcysteamine thioesters of novel precursor molecules to prepare narbonolide derivatives. See U.S. patent application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT publication Nos. WO 99/03986 and WO 97/02358, each of which is
15 incorporated herein by reference.

The recombinant DNA compounds of the invention that encode the second extender module of the narbonolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the narbonolide PKS second extender module is inserted into a DNA
20 compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the narbonolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the
25 second extender module of the narbonolide PKS is inserted into a DNA compound that comprises the coding sequences for the narbonolide PKS or a recombinant narbonolide PKS that produces a narbonolide derivative.

In another embodiment, a portion or all of the second extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid
30 module. In this embodiment, the invention provides, for example, replacing the malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or carboxyglycolyl CoA specific AT; deleting (or inactivating) the KR, the DH, or both the DH and KR; replacing the KR or the KR and DH with a KR, a KR and a DH, or a KR, DH, and ER; and/or inserting an ER. In

addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the narbonolide PKS, from a coding sequence for a PKS that produces a polyketide other than narbonolide, or from
5 chemical synthesis. The resulting heterologous second extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes narbonolide, a narbonolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the third extender module of the narbonolide PKS and the corresponding polypeptides encoded thereby are
10 useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the narbonolide PKS third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the third extender module of the narbonolide PKS or the latter is merely
15 added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the narbonolide PKS is inserted into a DNA compound that comprises coding sequences for the narbonolide PKS or a recombinant narbonolide PKS that produces a narbonolide derivative.

20 In another embodiment, a portion or all of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or carboxyglycolyl CoA specific AT; deleting the inactive KR; and/or inserting a KR, or a KR and DH, or a KR, DH, and ER. In addition,
25 the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the narbonolide PKS, from a gene for a PKS that produces a polyketide other than narbonolide, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized
30 in conjunction with a coding sequence for a PKS that synthesizes narbonolide, a narbonolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the fourth extender module of the narbonolide PKS and the corresponding polypeptides encoded thereby are

useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the narbonolide PKS fourth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either
5 replaced by that for the fourth extender module of the narbonolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fourth extender module of the narbonolide PKS is inserted into a DNA compound that comprises coding sequences for the narbonolide PKS or a recombinant narbonolide PKS that produces a
10 narbonolide derivative.

In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or carboxyglycolyl CoA specific AT; deleting
15 any one, two, or all three of the ER, DH, and KR; and/or replacing any one, two, or all three of the ER, DH, and KR with either a KR, a DH and KR, or a KR, DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the narbonolide PKS, from a
20 coding sequence for a PKS that produces a polyketide other than narbonolide, or from chemical synthesis. The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes narbonolide, a narbonolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the fifth extender
25 module of the narbonolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the narbonolide PKS fifth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either
30 replaced by that for the fifth extender module of the narbonolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fifth extender module of the narbonolide PKS is inserted into a DNA compound that comprises the

coding sequence for the narbonolide PKS or a recombinant narbonolide PKS that produces a narbonolide derivative.

In another embodiment, a portion or all of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or carboxyglycolyl CoA specific AT; deleting (or inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the narbonolide PKS, from a coding sequence for a PKS that produces a polyketide other than narbonolide, or from chemical synthesis. The resulting heterologous fifth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes narbonolide, a narbonolide derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the sixth extender module of the narbonolide PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the narbonolide PKS sixth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth extender module of the narbonolide PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the sixth extender module of the narbonolide PKS is inserted into a DNA compound that comprises the coding sequences for the narbonolide PKS or a recombinant narbonolide PKS that produces a narbonolide derivative.

In another embodiment, a portion or all of the sixth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or carboxyglycolyl CoA specific AT; and/or inserting a KR, a KR and DH, or a KR, DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the

heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the narbonolide PKS, from a coding sequence for a PKS that produces a polyketide other than narbonolide, or from chemical synthesis. The resulting heterologous sixth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes narbonolide, a narbonolide derivative, or another polyketide.

The sixth extender module of the narbonolide PKS is followed by a thioesterase domain. This domain is important in the cyclization of the polyketide and its cleavage from the PKS. The present invention provides recombinant DNA compounds that encode hybrid PKS enzymes in which the narbonolide PKS is fused to a heterologous thioesterase or a heterologous PKS is fused to the narbonolide synthase thioesterase. Thus, for example, a thioesterase domain coding sequence from another PKS gene can be inserted at the end of the sixth extender module coding sequence in recombinant DNA compounds of the invention. Recombinant DNA compounds encoding this thioesterase domain are therefore useful in constructing DNA compounds that encode the narbonolide PKS, a PKS that produces a narbonolide derivative, and a PKS that produces a polyketide other than narbonolide or a narbonolide derivative.

The following Table lists references describing illustrative PKS genes and corresponding enzymes that can be utilized in the construction of the recombinant hybrid PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing tailoring enzymes and corresponding genes that can be employed in accordance with the methods of the invention.

Avermectin

U.S. Pat. No. 5,252,474 to Merck.

MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemadectin.

MacNeil *et al.*, 1992, *Gene* 115: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

Candicidin (FR008)

Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172.

Epothilone

U.S. patent application Serial No. 60/130,560, filed 22 Apr. 1999, and Serial No. 60/122,620, filed 3 Mar. 1999.

Erythromycin

5 PCT Pub. No. WO 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, *Science* 252:675-9.

Cortes *et al.*, 8 Nov. 1990, *Nature* 348:176-8, An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*.
10

Glycosylation Enzymes

PCT Pat. App. Pub. No. WO 97/23630 to Abbott.

FK506

Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506, *Eur. J. Biochem.* 256: 528-534.
15

Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506, *Eur. J. Biochem.* 244: 74-80.

Methyltransferase

20 US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from *Streptomyces* MA6858. 31-O-desmethyl-FK506 methyltransferase.

Motamedi *et al.*, 1996, Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520, *J. Bacteriol.* 178: 5243-5248.

FK520

25 U.S. patent application Serial No. 60/123,800, filed 11 Mar. 1999.

Immunomycin

Nielsen *et al.*, 1991, *Biochem.* 30:5789-96.

Lovastatin

30 U.S. Pat. No. 5,744,350 to Merck.

Nemadectin

MacNeil *et al.*, 1993, *supra*.

Niddamycin

Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, *J. Bacteriol.* 179: 7515-7522.

Oleandomycin

- 5 Swan *et al.*, 1994, Characterization of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence, *Mol. Gen. Genet.* 242: 358-362.

- Olano *et al.*, 1998, Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring, *Mol. Gen. Genet.* 259(3): 299-308.

U.S. patent application Serial No. 60/120,254, filed 16 Feb. 1999, and Serial No. 60/106,100, filed 29 Oct. 1998.

Platenolide

EP Pat. App. Pub. No. 791,656 to Lilly.

15 **Pradimicin**

PCT Pat. Pub. No. WO 98/11230 to Bristol-Myers Squibb.

Rapamycin

Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, *Proc. Natl. Acad. Sci. USA* 92:7839-7843.

- 20 Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, *Gene* 169: 9-16.

Rifamycin

- August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amiclatopsis mediterranei* S669, *Chemistry & Biology*, 5(2): 69-79.

Soraphen

U.S. Pat. No. 5,716,849 to Novartis.

- Schupp *et al.*, 1995, *J. Bacteriology* 177: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

Spiramycin

U.S. Pat. No. 5,098,837 to Lilly.

Activator Gene

U.S. Pat. No. 5,514,544 to Lilly.

5 Tylosin

EP Pub. No. 791,655 to Lilly.

Kuhstoss *et al.*, 1996, *Gene* 183:231-6., Production of a novel polyketide through the construction of a hybrid polyketide synthase.

U.S. Pat. No. 5,876,991 to Lilly.

10 Tailoring enzymes

Merson-Davies and Cundliffe, 1994, *Mol. Microbiol.* 13: 349-355. Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome.

As the above Table illustrates, there is a wide variety of PKS genes that serve as readily available sources of DNA and sequence information for use in constructing the hybrid
15 PKS-encoding DNA compounds of the invention. Methods for constructing hybrid PKS-encoding DNA compounds are described without reference to the narbonolide PKS in U.S. Patent Nos. 5,672,491 and 5,712,146 and PCT publication No. WO 98/49315, each of which is incorporated herein by reference.

In constructing hybrid PKSs of the invention, certain general methods may be helpful.
20 For example, it is often beneficial to retain the framework of the module to be altered to make the hybrid PKS. Thus, if one desires to add DH and ER functionalities to a module, it is often preferred to replace the KR domain of the original module with a KR, DH, and ER domain-containing segment from another module, instead of merely inserting DH and ER domains. One can alter the stereochemical specificity of a module by replacement of the KS domain
25 with a KS domain from a module that specifies a different stereochemistry. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units" *Biochemistry* 38(5):1643-1651, incorporated herein by reference. One can alter the specificity of an AT domain by changing only a small segment of the domain. See Lau *et al.*, *supra*. One can also take advantage of
30 known linker regions in PKS proteins to link modules from two different PKSs to create a hybrid PKS. See Gokhale *et al.*, 16 Apr. 1999, "Dissecting and Exploiting Intermodular Communication in Polyketide Synthases", *Science* 284: 482-485, incorporated herein by reference.

The hybrid PKS-encoding DNA compounds of the invention can be and often are hybrids of more than two PKS genes. Even where only two genes are used, there are often two or more modules in the hybrid gene in which all or part of the module is derived from a second (or third) PKS gene. Thus, as one illustrative example, the invention provides a hybrid
5 narbonolide PKS that contains the naturally occurring loading module and thioesterase domain as well as extender modules one, two, four, and six of the narbonolide PKS and further contains hybrid or heterologous extender modules three and five. Hybrid or heterologous extender modules three and five contain AT domains specific for malonyl CoA and derived from, for example, the rapamycin PKS genes.

10 To construct a hybrid PKS or narbonolide derivative PKS of the invention, one can employ a technique, described in PCT Pub. No. WO 98/27203, which is incorporated herein by reference, in which the large PKS gene cluster is divided into two or more, typically three, segments, and each segment is placed on a separate expression vector. In this manner, each of the segments of the gene can be altered, and various altered segments can be combined in a
15 single host cell to provide a recombinant PKS gene of the invention. This technique makes more efficient the construction of large libraries of recombinant PKS genes, vectors for expressing those genes, and host cells comprising those vectors.

Included in the definition of "hybrid" are PKS where alterations (including deletions, insertions and substitutions) are made directly using the narbonolide PKS as a substrate.

20 The invention also provides libraries of PKS genes, PKS proteins, and ultimately, of polyketides, that are constructed by generating modifications in the narbonolide PKS so that the protein complexes produced have altered activities in one or more respects and thus produce polyketides other than the natural product of the PKS. Novel polyketides may thus be prepared, or polyketides in general prepared more readily, using this method. By providing
25 a large number of different genes or gene clusters derived from a naturally occurring PKS gene cluster, each of which has been modified in a different way from the native cluster, an effectively combinatorial library of polyketides can be produced as a result of the multiple variations in these activities. As will be further described below, the metes and bounds of this embodiment of the invention can be described on both the protein level and the encoding
30 nucleotide sequence level.

As described above, a modular PKS "derived from" the narbonolide or other naturally occurring PKS is a subset of the "hybrid" PKS family and includes a modular PKS (or its corresponding encoding gene(s)) that retains the scaffolding of the utilized portion of the

naturally occurring gene. Not all modules need be included in the constructs. On the constant scaffold, at least one enzymatic activity is mutated, deleted, replaced, or inserted so as to alter the activity of the resulting PKS relative to the original PKS. Alteration results when these activities are deleted or are replaced by a different version of the activity, or simply mutated

5 in such a way that a polyketide other than the natural product results from these collective activities. This occurs because there has been a resulting alteration of the starter unit and/or extender unit, and/or stereochemistry, and/or chain length or cyclization, and/or reductive or dehydration cycle outcome at a corresponding position in the product polyketide. Where a deleted activity is replaced, the origin of the replacement activity may come from a

10 corresponding activity in a different naturally occurring PKS or from a different region of the narbonolide PKS. Any or all of the narbonolide PKS genes may be included in the derivative or portions of any of these may be included, but the scaffolding of the PKS protein is retained in whatever derivative is constructed. The derivative preferably contains a thioesterase activity from the narbonolide or another PKS.

15 In summary, a PKS "derived from" the narbonolide PKS includes a PKS that contains the scaffolding of all or a portion of the narbonolide PKS. The derived PKS also contains at least two extender modules that are functional, preferably three extender modules, and more preferably four or more extender modules, and most preferably six extender modules. The derived PKS also contains mutations, deletions, insertions, or replacements of one or more of

20 the activities of the functional modules of the narbonolide PKS so that the nature of the resulting polyketide is altered. This definition applies both at the protein and DNA sequence levels. Particular preferred embodiments include those wherein a KS, AT, KR, DH, or ER has been deleted or replaced by a version of the activity from a different PKS or from another location within the same PKS. Also preferred are derivatives where at least one non-

25 condensation cycle enzymatic activity (KR, DH, or ER) has been deleted or added or wherein any of these activities has been mutated so as to change the structure of the polyketide synthesized by the PKS.

Conversely, also included within the definition of a PKS derived from the narbonolide PKS are functional PKS modules or their encoding genes wherein at least one portion,

30 preferably two portions, of the narbonolide PKS activities have been inserted. Exemplary is the use of the narbonolide AT for extender module 2 which accepts a malonyl CoA extender unit rather than methylmalonyl CoA to replace a methylmalonyl specific AT in a PKS. Other examples include insertion of portions of non-condensation cycle enzymatic activities or

other regions of narbonolide synthase activity into a heterologous PKS. Again, the derived from definition applies to the PKS at both the genetic and protein levels.

Thus, there are at least five degrees of freedom for constructing a hybrid PKS in terms of the polyketide that will be produced. First, the polyketide chain length is determined by the number of modules in the PKS. Second, the nature of the carbon skeleton of the PKS is determined by the specificities of the acyl transferases that determine the nature of the extender units at each position, e.g., malonyl, methylmalonyl, ethylmalonyl, or other substituted malonyl. Third, the loading module specificity also has an effect on the resulting carbon skeleton of the polyketide. The loading module may use a different starter unit, such as acetyl, butyryl, and the like. As noted above and in the examples below, another method for varying loading module specificity involves inactivating the KS activity in extender module 1 (KS1) and providing alternative substrates, called diketides that are chemically synthesized analogs of extender module 1 diketide products, for extender module 2. This approach was illustrated in PCT publication Nos. WO 97/02358 and WO 99/03986, incorporated herein by reference, wherein the KS1 activity was inactivated through mutation. Fourth, the oxidation state at various positions of the polyketide will be determined by the dehydratase and reductase portions of the modules. This will determine the presence and location of ketone and alcohol moieties and C-C double bonds or C-C single bonds in the polyketide. Finally, the stereochemistry of the resulting polyketide is a function of three aspects of the synthase. The first aspect is related to the AT/KS specificity associated with substituted malonyls as extender units, which affects stereochemistry only when the reductive cycle is missing or when it contains only a ketoreductase, as the dehydratase would abolish chirality. Second, the specificity of the ketoreductase may determine the chirality of any beta-OH. Finally, the enoylreductase specificity for substituted malonyls as extender units may influence the result when there is a complete KR/DH/ER available.

Thus, the modular PKS systems, and in particular, the narbonolide PKS system, permit a wide range of polyketides to be synthesized. As compared to the aromatic PKS systems, a wider range of starter units including aliphatic monomers (acetyl, propionyl, butyryl, isovaleryl, etc.), aromatics (aminohydroxybenzoyl), alicyclics (cyclohexanoyl), and heterocyclics (thiazolyl) are found in various macrocyclic polyketides. Recent studies have shown that modular PKSs have relaxed specificity for their starter units (Kao *et al.*, 1994, *Science, supra*). Modular PKSs also exhibit considerable variety with regard to the choice of extender units in each condensation cycle. The degree of beta-ketoreduction following a

condensation reaction has also been shown to be altered by genetic manipulation (Donadio *et al.*, 1991, *Science, supra*; Donadio *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90: 7119-7123).

Likewise, the size of the polyketide product can be varied by designing mutants with the appropriate number of modules (Kao *et al.*, 1994, *J. Am. Chem. Soc.* 116:11612-11613).

- 5 Lastly, these enzymes are particularly well known for generating an impressive range of asymmetric centers in their products in a highly controlled manner. The polyketides and antibiotics produced by the methods of the invention are typically single stereoisomeric forms. Although the compounds of the invention can occur as mixtures of stereoisomers, it may be beneficial in some instances to generate individual stereoisomers. Thus, the
- 10 combinatorial potential within modular PKS pathways based on any naturally occurring modular, such as the narbonolide, PKS scaffold is virtually unlimited.

- The combinatorial potential is increased even further when one considers that mutations in DNA encoding a polypeptide can be used to introduce, alter, or delete an activity in the encoded polypeptide. Mutations can be made to the native sequences using
- 15 conventional techniques. The substrates for mutation can be an entire cluster of genes or only one or two of them; the substrate for mutation may also be portions of one or more of these genes. Techniques for mutation include preparing synthetic oligonucleotides including the mutations and inserting the mutated sequence into the gene encoding a PKS subunit using restriction endonuclease digestion. See, e.g., Kunkel, 1985, *Proc. Natl. Acad. Sci. USA* 82:
- 20 448; Geisselsoder *et al.*, 1987, *BioTechniques* 5:786. Alternatively, the mutations can be effected using a mismatched primer (generally 10-20 nucleotides in length) that hybridizes to the native nucleotide sequence, at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within relatively narrow limits and by keeping the mutant base centrally located.
- 25 See Zoller and Smith, 1983, *Methods Enzymol.* 100:468. Primer extension is effected using DNA polymerase, the product cloned, and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Identification can be accomplished using the mutant primer as a hybridization probe. The technique is also applicable for generating multiple point mutations. See, e.g., Dalbie-McFarland *et al.*, 1982, *Proc. Natl. Acad. Sci.*
- 30 *USA* 79: 6409. PCR mutagenesis can also be used to effect the desired mutations.

Random mutagenesis of selected portions of the nucleotide sequences encoding enzymatic activities can also be accomplished by several different techniques known in the art, e.g., by inserting an oligonucleotide linker randomly into a plasmid, by irradiation with

X-rays or ultraviolet light, by incorporating incorrect nucleotides during *in vitro* DNA synthesis, by error-prone PCR mutagenesis, by preparing synthetic mutants, or by damaging plasmid DNA *in vitro* with chemicals. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, nitrosoguanidine, hydroxylamine, agents which damage or remove
5 bases thereby preventing normal base-pairing such as hydrazine or formic acid, analogues of nucleotide precursors such as 5-bromouracil, 2-aminopurine, or acridine intercalating agents such as proflavine, acriflavine, quinacrine, and the like. Generally, plasmid DNA or DNA fragments are treated with chemicals, transformed into *E. coli* and propagated as a pool or library of mutant plasmids.

10 In constructing a hybrid PKS of the invention, regions encoding enzymatic activity, i.e., regions encoding corresponding activities from different PKS synthases or from different locations in the same PKS, can be recovered, for example, using PCR techniques with appropriate primers. By "corresponding" activity encoding regions is meant those regions encoding the same general type of activity. For example, a KR activity encoded at one
15 location of a gene cluster "corresponds" to a KR encoding activity in another location in the gene cluster or in a different gene cluster. Similarly, a complete reductase cycle could be considered corresponding. For example, KR/DH/ER corresponds to KR alone.

If replacement of a particular target region in a host PKS is to be made, this replacement can be conducted *in vitro* using suitable restriction enzymes. The replacement
20 can also be effected *in vivo* using recombinant techniques involving homologous sequences framing the replacement gene in a donor plasmid and a receptor region in a recipient plasmid. Such systems, advantageously involving plasmids of differing temperature sensitivities are described, for example, in PCT publication No. WO 96/40968, incorporated herein by reference. The vectors used to perform the various operations to replace the enzymatic
25 activity in the host PKS genes or to support mutations in these regions of the host PKS genes can be chosen to contain control sequences operably linked to the resulting coding sequences in a manner such that expression of the coding sequences can be effected in an appropriate host.

However, simple cloning vectors may be used as well. If the cloning vectors
30 employed to obtain PKS genes encoding derived PKS lack control sequences for expression operably linked to the encoding nucleotide sequences, the nucleotide sequences are inserted into appropriate expression vectors. This need not be done individually, but a pool of isolated encoding nucleotide sequences can be inserted into expression vectors, the resulting vectors

transformed or transfected into host cells, and the resulting cells plated out into individual colonies.

The various PKS nucleotide sequences can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements, or under the control of, e.g., a
5 single promoter. The PKS subunit encoding regions can include flanking restriction sites to allow for the easy deletion and insertion of other PKS subunit encoding sequences so that hybrid PKSs can be generated. The design of such unique restriction sites is known to those of skill in the art and can be accomplished using the techniques described above, such as site-directed mutagenesis and PCR.

10 The expression vectors containing nucleotide sequences encoding a variety of PKS enzymes for the production of different polyketides are then transformed into the appropriate host cells to construct the library. In one straightforward approach, a mixture of such vectors is transformed into the selected host cells and the resulting cells plated into individual colonies and selected to identify successful transformants. Each individual colony has the
15 ability to produce a particular PKS synthase and ultimately a particular polyketide. Typically, there will be duplications in some, most, or all of the colonies; the subset of the transformed colonies that contains a different PKS in each member colony can be considered the library. Alternatively, the expression vectors can be used individually to transform hosts, which transformed hosts are then assembled into a library. A variety of strategies are available to
20 obtain a multiplicity of colonies each containing a PKS gene cluster derived from the naturally occurring host gene cluster so that each colony in the library produces a different PKS and ultimately a different polyketide. The number of different polyketides that are produced by the library is typically at least four, more typically at least ten, and preferably at least 20, and more preferably at least 50, reflecting similar numbers of different altered PKS
25 gene clusters and PKS gene products. The number of members in the library is arbitrarily chosen; however, the degrees of freedom outlined above with respect to the variation of starter, extender units, stereochemistry, oxidation state, and chain length is quite large.

Methods for introducing the recombinant vectors of the invention into suitable hosts are known to those of skill in the art and typically include the use of CaCl_2 or agents such as
30 other divalent cations, lipofection, DMSO, protoplast transformation, infection, transfection, and electroporation. The polyketide producing colonies can be identified and isolated using known techniques and the produced polyketides further characterized. The polyketides

produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity.

The libraries of the invention can thus be considered at four levels: (1) a multiplicity of colonies each with a different PKS encoding sequence; (2) colonies that contain the
5 proteins that are members of the PKS library produced by the coding sequences; (3) the polyketides produced; and (4) antibiotics or compounds with other desired activities derived from the polyketides. Of course, combination libraries can also be constructed wherein members of a library derived, for example, from the narbonolide PKS can be considered as a part of the same library as those derived from, for example, the rapamycin PKS or DEBS.

10 Colonies in the library are induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of polyketides. The polyketides secreted into the media can be screened for binding to desired targets, such as receptors, signaling proteins, and the like. The supernatants per se can be used for screening, or partial or complete purification of the polyketides can first be effected. Typically, such screening
15 methods involve detecting the binding of each member of the library to receptor or other target ligand. Binding can be detected either directly or through a competition assay. Means to screen such libraries for binding are well known in the art. Alternatively, individual polyketide members of the library can be tested against a desired target. In this event, screens wherein the biological response of the target is measured can more readily be included.
20 Antibiotic activity can be verified using typical screening assays such as those set forth in Lehrer *et al.*, 1991, *J. Immunol. Meth.* 137:167-173, incorporated herein by reference, and in the examples below.

The invention provides methods for the preparation of a large number of polyketides. These polyketides are useful intermediates in formation of compounds with antibiotic or
25 other activity through hydroxylation and glycosylation reactions as described above. In general, the polyketide products of the PKS must be further modified, typically by hydroxylation and glycosylation, to exhibit antibiotic activity. Hydroxylation results in the novel polyketides of the invention that contain hydroxyl groups at C6, which can be accomplished using the hydroxylase encoded by the *eryF* gene, and/or C12, which can be
30 accomplished using the hydroxylase encoded by the *picK* or *eryK* gene. The presence of hydroxyl groups at these positions can enhance the antibiotic activity of the resulting compound relative to its unhydroxylated counterpart.

Glycosylation is important in conferring antibiotic activity to a polyketide as well. Methods for glycosylating the polyketides are generally known in the art; the glycosylation may be effected intracellularly by providing the appropriate glycosylation enzymes or may be effected *in vitro* using chemical synthetic means as described herein and in PCT publication
5 No. WO 98/49315, incorporated herein by reference. Preferably, glycosylation with desosamine is effected in accordance with the methods of the invention in recombinant host cells provided by the invention. In general, the approaches to effecting glycosylation mirror those described above with respect to hydroxylation. The purified enzymes, isolated from native sources or recombinantly produced may be used *in vitro*. Alternatively and as noted,
10 glycosylation may be effected intracellularly using endogenous or recombinantly produced intracellular glycosylases. In addition, synthetic chemical methods may be employed.

The antibiotic modular polyketides may contain any of a number of different sugars, although D-desosamine, or a close analog thereof, is most common. Erythromycin, picromycin, narbomycin and methymycin contain desosamine. Erythromycin also contains L-
15 cladinose (3-O-methyl mycarose). Tylosin contains mycaminoses (4-hydroxy desosamine), mycarose and 6-deoxy-D-allose. 2-acetyl-1-bromodesosamine has been used as a donor to glycosylate polyketides by Masamune *et al.*, 1975, *J. Am. Chem. Soc.* 97: 3512-3513. Other, apparently more stable donors include glycosyl fluorides, thioglycosides, and trichloroacetimidates; see Woodward *et al.*, 1981, *J. Am. Chem. Soc.* 103: 3215; Martin *et al.*,
20 1997, *J. Am. Chem. Soc.* 119: 3193; Toshima *et al.*, 1995, *J. Am. Chem. Soc.* 117: 3717; Matsumoto *et al.*, 1988, *Tetrahedron Lett.* 29: 3575. Glycosylation can also be effected using the polyketide aglycones as starting materials and using *Saccharopolyspora erythraea* or *Streptomyces venezuelae* to make the conversion, preferably using mutants unable to synthesize macrolides.

25 To provide an illustrative hybrid PKS of the invention as well as an expression vector for that hybrid PKS and host cells comprising the vector and producing the hybrid polyketide, a portion of the narbonolide PKS gene was fused to the DEBS genes. This construct also allowed the examination of whether the TE domain of the narbonolide PKS (*pikTE*) could promote formation of 12-membered lactones in the context of a different PKS. A construct
30 was generated, plasmid pKOS039-18, in which the *pikTE* ORF was fused with the DEBS genes in place of the DEBS TE ORF (see Figure 5). To allow the TE to distinguish between substrates most closely resembling those generated by the narbonolide PKS, the fusion junction was chosen between the AT and ACP to eliminate ketoreductase activity in DEBS

extender module 6 (KR⁶). This results in a hybrid PKS that presents the TE with a β -ketone heptaketide intermediate and a β -(S)-hydroxy hexaketide intermediate to cyclize, as in narbonolide and 10-deoxymethynolide biosynthesis.

Analysis of this construct indicated the production of the 14-membered ketolide 3,6-dideoxy-3-oxo-erythronolide B (Figure 5, compound 6). Extracts were analyzed by LC/MS. The identity of compound 6 was verified by comparison to a previously authenticated sample (see PCT publication No. WO 98/49315, incorporated herein by reference). The predicted 12-membered macrolactone, (8R,9S)-8,9-dihydro-8-methyl-9-hydroxy-10-deoxymethynolide (see Kao *et al.* *J. Am. Chem. Soc.* (1995) 117:9105-9106 incorporated herein by reference) was not detected. Because the 12-membered intermediate can be formed by other recombinant PKS enzymes, see Kao *et al.*, 1995, *supra*, the PIC TE domain appears incapable of forcing premature cyclization of the hexaketide intermediate generated by DEBS. This result, along with others reported herein, suggests that protein interactions between the narbonolide PKS modules play a role in formation of the 12 and 14-membered macrolides.

The above example illustrates also how engineered PKSs can be improved for production of novel compounds. Compound 6 was originally produced by deletion of the KR⁶ domain in DEBS to create a 3-ketolide producing PKS (see U.S. patent application Serial No. 09/073,538, filed 6 May 1998, and PCT publication No. WO 98/49315, each of which is incorporated herein by reference). Although the desired molecule was made, purification of compound 6 from this strain was hampered by the presence of 2-desmethyl ketolides that could not be easily separated. Extracts from *Streptomyces lividans* K4-114/pKOS039-18, however, do not contain the 2-desmethyl compounds, greatly simplifying purification. Thus, the invention provides a useful method of producing such compounds. The ability to combine the narbonolide PKS with DEBS and other modular PKSs provides a significant advantage in the production of macrolide antibiotics.

Two other hybrid PKSs of the invention were constructed that yield this same compound. These constructs also illustrate the method of the invention in which hybrid PKSs are constructed at the protein, as opposed to the module, level. Thus, the invention provides a method for constructing a hybrid PKS which comprises the coexpression of at least one gene from a first modular PKS gene cluster in a host cell that also expresses at least one gene from a second PKS gene cluster. The invention also provides novel hybrid PKS enzymes prepared in accordance with the method. This method is not limited to hybrid PKS enzymes composed

of at least one narbonolide PKS gene, although such constructs are illustrative and preferred. Moreover, the hybrid PKS enzymes are not limited to hybrids composed of unmodified proteins; as illustrated below, at least one of the genes can optionally be a hybrid PKS gene.

In the first construct, the *eryAI* and *eryAII* genes were coexpressed with *picAIV* and a
5 gene encoding a hybrid extender module 5 composed of the KS and AT domains of extender module 5 of DEBS3 and the KR and ACP domains of extender module 5 of the narbonolide PKS. In the second construct, the *picAIV* coding sequence was fused to the hybrid extender module 5 coding sequence used in the first construct to yield a single protein. Each of these constructs produced 3-deoxy-3-oxo-6-deoxyerythronolide B. In a third construct, the coding
10 sequence for extender module 5 of DEBS3 was fused to the *picAIV* coding sequence, but the levels of product produced were below the detection limits of the assay.

A variant of the first construct hybrid PKS was constructed that contained an inactivated DEBS1 extender module 1 KS domain. When host cells containing the resultant hybrid PKS were supplied the appropriate diketide precursor, the desired 13-desethyl-13-
15 propyl compounds were obtained, as described in the examples below.

Other illustrative hybrid PKSs of the invention were made by coexpressing the *picAI* and *picAII* genes with genes encoding DEBS3 or DEBS3 variants. These constructs illustrate the method of the invention in which a hybrid PKS is produced from coexpression of PKS genes unmodified at the modular or domain level. In the first construct, the *eryAIII* gene was
20 coexpressed with the *picAI* and *picAII* genes, and the hybrid PKS produced 10-desmethyl-10,11-anhydro-6-deoxyerythronolide B in *Streptomyces lividans*. Such a hybrid PKS could also be constructed in accordance with the method of the invention by transformation of *S. venezuelae* with an expression vector that produces the *eryAIII* gene product, DEBS3. In a preferred embodiment, the *S. venezuelae* host cell has been modified to inactivate the *picAIII*
25 gene.

In the second construct, the DEBS3 gene was a variant that had an inactive KR in extender module 5. The hybrid PKS produced 5,6-dideoxy-5-oxo-10-desmethyl-10,11-anhydroerythronolide B in *Streptomyces lividans*.

In the third construct, the DEBS3 gene was a variant in which the KR domain of
30 extender module 5 was replaced by the DH and KR domains of extender module 4 of the rapamycin PKS. This construct produced 5,6-dideoxy-5-oxo-10-desmethyl-10,11-anhydroerythronolide B and 5,6-dideoxy-4,5-anhydro-10-desmethyl-10,11-

anhydroerythronolide B in *Streptomyces lividans*, indicating that the rapamycin DH and KR domains functioned only inefficiently in this construct.

In the fourth construct, the DEBS3 gene was a variant in which the KR domain of extender module 5 was replaced by the DH, KR, and ER domains of extender module 1 of the rapamycin PKS. This construct produced 5,6-dideoxy-5-oxo-10-desmethyl-10,11-anhydroerythronolide B as well as 5,6-dideoxy-10-desmethyl-10,11-anhydroerythronolide B in *Streptomyces lividans*, indicating that the rapamycin DH, KR, and ER domains functioned only inefficiently in this construct.

In the fifth construct, the DEBS3 gene was a variant in which the KR domain of extender module 6 was replaced by the DH and KR domains of extender module 4 of the rapamycin PKS. This construct produced 3,6-dideoxy-2,3-anhydro-10-desmethyl-10,11-anhydroerythronolide B in *Streptomyces lividans*.

In the sixth construct, the DEBS3 gene was a variant in which the AT domain of extender module 6 was replaced by the AT domain of extender module 2 of the rapamycin PKS. This construct produced 2,10-didesmethyl-10,11-anhydro-6-deoxyerythronolide B in *Streptomyces lividans*.

These hybrid PKSs illustrate the wide variety of polyketides that can be produced by the methods and compounds of the invention. These polyketides are useful as antibiotics and as intermediates in the synthesis of other useful compounds, as described in the following section.

Section VI: Compounds

The methods and recombinant DNA compounds of the invention are useful in the production of polyketides. In one important aspect, the invention provides methods for making ketolides, polyketide compounds with significant antibiotic activity. See Griesgraber *et al.*, 1996, *J. Antibiot.* 49: 465-477, incorporated herein by reference. Most if not all of the ketolides prepared to date are synthesized using erythromycin A, a derivative of 6-dEB, as an intermediate. While the invention provides hybrid PKSs that produce a polyketide different in structure from 6-dEB, the invention also provides methods for making intermediates useful in preparing traditional, 6-dEB-derived ketolide compounds.

Because 6-dEB in part differs from narbonolide in that it comprises a 10-methyl group, the novel hybrid PKS genes of the invention based on the narbonolide PKS provide many novel ketolides that differ from the known ketolides only in that they lack a 10-methyl

group. Thus, the invention provides the 10-desmethyl analogues of the ketolides and intermediates and precursor compounds described in, for example, Griesgraber *et al.*, *supra*; Agouridas *et al.*, 1998, *J. Med. Chem.* 41: 4080-4100, U.S. Patent Nos. 5,770,579; 5,760,233; 5,750,510; 5,747,467; 5,747,466; 5,656,607; 5,635,485; 5,614,614; 5,556,118; 5,543,400; 5,527,780; 5,444,051; 5,439,890; 5,439,889; and PCT publication Nos. WO 98/09978 and WO 98/28316, each of which is incorporated herein by reference. Because the invention also provides hybrid PKS genes that include a methylmalonyl-specific AT domain in extender module 2 of the narbonolide PKS, the invention also provides hybrid PKS that can be used to produce the 10-methyl-containing ketolides known in the art.

10 Thus, a hybrid PKS of the invention that produces 10-methyl narbonolide is constructed by substituting the malonyl-specific AT domain of the narbonolide PKS extender module 2 with a methylmalonyl specific AT domain from a heterologous PKS. A hybrid narbonolide PKS in which the AT of extender module 2 was replaced with the AT from DEBS extender module 2 was constructed using boundaries described in PCT publication
15 No. WO 98/49315, incorporated herein by reference. However, when the hybrid PKS expression vector was introduced into *Streptomyces venezuelae*, detectable quantities of 10-methyl picromycin were not produced. Thus, to construct such a hybrid PKS of the invention, an AT domain from a module other than DEBS extender module 2 is preferred. One could also employ DEBS extender module 2 or another methylmalonyl specific AT but utilize
20 instead different boundaries than those used for the substitution described above. In addition, one can construct such a hybrid PKS by substituting, in addition to the AT domain, additional extender module 2 domains, including the KS, the KR, and the DH, and/or additional extender module 3 domains.

 Although modification of extender module 2 of the narbonolide PKS is required, the
25 extent of hybrid modules engineered need not be limited to module 2 to make 10-methyl narbonolide. For example, substitution of the KS domain of extender module 3 of the narbonolide PKS with a heterologous domain or module can result in more efficient processing of the intermediate generated by the hybrid extender module 2. Likewise, a heterologous TE domain may be more efficient in cyclizing 10-methyl narbonolide.

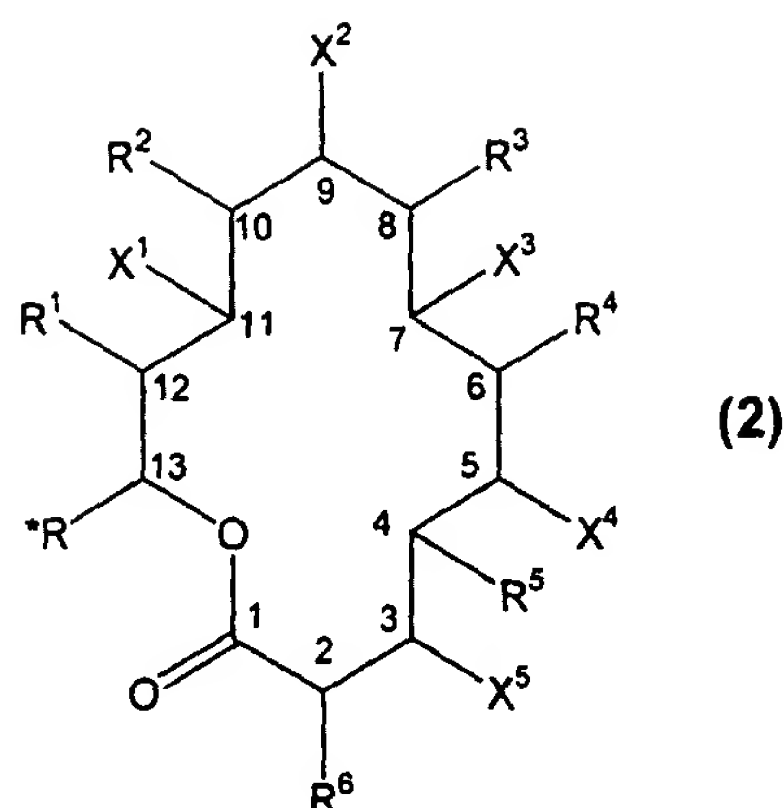
30 Substitution of the entire extender module 2 of the narbonolide PKS with a module encoding the correct enzymatic activities, i.e., a KS, a methylmalonyl specific AT, a KR, a DH, and an ACP, can also be used to create a hybrid PKS of the invention that produces a 10-methyl ketolide. Modules useful for such whole module replacements include extender

modules 4 and 10 from the rapamycin PKS, extender modules 1 and 5 from the FK506 PKS, extender module 2 of the tylosin PKS, and extender module 4 of the rifamycin PKS. Thus, the invention provides many different hybrid PKSs that can be constructed starting from the narbonolide PKS that can be used to produce 10-methyl narbonolide. While 10-methyl
5 narbonolide is referred to in describing these hybrid PKSs, those of skill recognize that the invention also therefore provides the corresponding derivatives produced by glycosylation and hydroxylation. For example, if the hybrid PKS is expressed in *Streptomyces narbonensis* or *S. venezuelae*, the compounds produced are 10-methyl narbomycin and picromycin, respectively. Alternatively, the PKS can be expressed in a host cell transformed with the
10 vectors of the invention that encode the desosamine biosynthesis and desosaminyl transferase and *picK* hydroxylase genes.

Other important compounds provided by the invention are the 6-hydroxy ketolides. These compounds include 3-deoxy-3-oxo erythronolide B, 6-hydroxy narbonolide, and 6-hydroxy-10-methyl narbonolide. In the examples below, the invention provides a method for
15 utilizing EryF to hydroxylate 3-ketolides that is applicable for the production of any 6-hydroxy-3-ketolide.

Thus, the hybrid PKS genes of the invention can be expressed in a host cell that contains the desosamine biosynthetic genes and desosaminyl transferase gene as well as the required hydroxylase gene(s), which may be either *picK* (for the C12 position) or *eryK* (for
20 the C12 position) and/or *eryF* (for the C6 position). The resulting compounds have antibiotic activity but can be further modified, as described in the patent publications referenced above, to yield a desired compound with improved or otherwise desired properties. Alternatively, the aglycone compounds can be produced in the recombinant host cell, and the desired glycosylation and hydroxylation steps carried out *in vitro* or *in vivo*, in the latter case by
25 supplying the converting cell with the aglycone.

The compounds of the invention are thus optionally glycosylated forms of the polyketide set forth in formula (2) below which are hydroxylated at either the C6 or the C12 or both. The compounds of formula (2) can be prepared using the loading and the six
30 extender modules of a modular PKS, modified or prepared in hybrid form as herein described. These polyketides have the formula:



including the glycosylated and isolated stereoisomeric forms thereof;

wherein R* is a straight chain, branched or cyclic, saturated or unsaturated substituted or unsubstituted hydrocarbyl of 1-15C;

5 each of R¹-R⁶ is independently H or alkyl (1-4C) wherein any alkyl at R¹ may optionally be substituted;

each of X¹-X⁵ is independently two H, H and OH, or =O; or

each of X¹-X⁵ is independently H and the compound of formula (2) contains a double-bond in the ring adjacent to the position of said X at 2-3, 4-5, 6-7, 8-9 and/or 10-11;

10 with the proviso that:

at least two of R¹-R⁶ are alkyl (1-4C).

Preferred compounds comprising formula 2 are those wherein at least three of R¹-R⁵ are alkyl (1-4C), preferably methyl or ethyl; more preferably wherein at least four of R¹-R⁵ are alkyl (1-4C), preferably methyl or ethyl. Also preferred are those wherein X² is two H, 15 =O, or H and OH, and/or X³ is H, and/or X¹ is OH and/or X⁴ is OH and/or X⁵ is OH. Also preferred are compounds with variable R* when R¹-R⁵ is methyl, X² is =O, and X¹, X⁴ and X⁵ are OH. The glycosylated forms of the foregoing are also preferred.

The invention also provides the 12-membered macrolides corresponding to the compounds above but produced from a narbonolide-derived PKS lacking extender modules 5 20 and 6 of the narbonolide PKS.

The compounds of the invention can be produced by growing and fermenting the host cells of the invention under conditions known in the art for the production of other polyketides. The compounds of the invention can be isolated from the fermentation broths of these cultured cells and purified by standard procedures. The compounds can be readily

formulated to provide the pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation will contain one or more of the compounds of the invention as an active ingredient in admixture with an organic or inorganic carrier or excipient suitable for external, enteral, or parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use.

The carriers which can be used include water, glucose, lactose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquefied form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl methylcellulose essentially as described in U.S. Patent No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as described in EPO patent publication No. 428,169, incorporated herein by reference.

Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, *Transplantation Proceedings XLX*, Supp. 6: 17-22, incorporated herein by reference. Dosage forms for external application may be prepared essentially as described in EPO patent publication No. 423,714, incorporated herein by reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

For the treatment of conditions and diseases caused by infection, a compound of the invention may be administered orally, topically, parenterally, by inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intramuscular, and intrasternal injection or infusion techniques.

Dosage levels of the compounds of the invention are of the order from about 0.01 mg to about 50 mg per kilogram of body weight per day, preferably from about 0.1 mg to about 10 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the invention may be administered on an intermittent basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for oral administration to humans may contain from 0.5 mg to 5 gm of active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds of the invention may be formulated within the range of, for example, 0.00001% to 60% by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight.

It will be understood, however, that the specific dose level for any particular patient will depend on a variety of factors. These factors include the activity of the specific compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the invention and shall not be construed as being a limitation on the scope of the invention or claims.

20

Example 1

General Methodology

Bacterial strains, plasmids, and culture conditions. *Streptomyces coelicolor* CH999 described in WO 95/08548, published 30 March 1995, or *S. lividans* K4-114, described in Ziermann and Betlach, Jan. '99, Recombinant Polyketide Synthesis in *Streptomyces*: Engineering of Improved Host Strains, BioTechniques 26:106-110, incorporated herein by reference, was used as an expression host. DNA manipulations were performed in *Escherichia coli* XL1-Blue, available from Stratagene. *E. coli* MC1061 is also suitable for use as a host for plasmid manipulation. Plasmids were passaged through *E. coli* ET12567 (dam dcm hsdS Cmr) (MacNeil, 1988, *J. Bacteriol.* 170: 5607, incorporated herein by reference) to generate unmethylated DNA prior to transformation of *S. coelicolor*. *E. coli* strains were grown under standard conditions. *S. coelicolor* strains were grown on R2YE agar

plates (Hopwood *et al.*, *Genetic manipulation of Streptomyces. A laboratory manual*. The John Innes Foundation: Norwich, 1985, incorporated herein by reference).

Many of the expression vectors of the invention illustrated in the examples are derived from plasmid pRM5, described in WO 95/08548, incorporated herein by reference.

5 This plasmid includes a colEI replicon, an appropriately truncated SCP2* *Streptomyces* replicon, two act-promoters to allow for bidirectional cloning, the gene encoding the *actII-ORF4* activator which induces transcription from act promoters during the transition from growth phase to stationary phase, and appropriate marker genes. Engineered restriction sites in the plasmid facilitate the combinatorial construction of PKS gene clusters starting from

10 cassettes encoding individual domains of naturally occurring PKSs. When plasmid pRM5 is used for expression of a PKS, all relevant biosynthetic genes can be plasmid-borne and therefore amenable to facile manipulation and mutagenesis in *E. coli*. This plasmid is also suitable for use in *Streptomyces* host cells. *Streptomyces* is genetically and physiologically well-characterized and expresses the ancillary activities required for *in vivo* production of

15 most polyketides. Plasmid pRM5 utilizes the act promoter for PKS gene expression, so polyketides are produced in a secondary metabolite-like manner, thereby alleviating the toxic effects of synthesizing potentially bioactive compounds *in vivo*.

Manipulation of DNA and organisms. Polymerase chain reaction (PCR) was performed using Pfu polymerase (Stratagene; Taq polymerase from Perkin Elmer Cetus can

20 also be used) under conditions recommended by the enzyme manufacturer. Standard *in vitro* techniques were used for DNA manipulations (Sambrook *et al.* *Molecular Cloning: A Laboratory Manual* (Current Edition)). *E. coli* was transformed using standard calcium chloride-based methods; a Bio-Rad *E. coli* pulsing apparatus and protocols provided by Bio-Rad could also be used. *S. coelicolor* was transformed by standard procedures (Hopwood *et*

25 *al.* *Genetic manipulation of Streptomyces. A laboratory manual*. The John Innes Foundation: Norwich, 1985), and depending on what selectable marker was employed, transformants were selected using 1 mL of a 1.5 mg/mL thiostrepton overlay, 1 mL of a 2 mg/mL apramycin overlay, or both.

30

Example 2

Cloning of the Picromycin Biosynthetic Gene Cluster from *Streptomyces venezuelae*

Genomic DNA (100 µg) isolated from *Streptomyces venezuelae* ATCC15439 using standard procedures was partially digested with Sau3AI endonuclease to generate fragments

~40 kbp in length. SuperCosI (Stratagene) DNA cosmid arms were prepared as directed by the manufacturer. A cosmid library was prepared by ligating 2.5 µg of the digested genomic DNA with 1.5 µg of cosmid arms in a 20 µL reaction. One microliter of the ligation mixture was propagated in *E. coli* XL1-Blue MR (Stratagene) using a GigapackIII XL packaging
5 extract kit (Stratagene). The resulting library of ~3000 colonies was plated on a 10x150 mm agar plate and replicated to a nylon membrane.

The library was initially screened by direct colony hybridization with a DNA probe specific for ketosynthase domain coding sequences of PKS genes. Colonies were alkaline lysed, and the DNA was crosslinked to the membrane using UV irradiation. After overnight
10 incubation with the probe at 42°C, the membrane was washed twice at 25°C in 2xSSC buffer + 0.1% SDS for 15 minutes, followed by two 15 minute washes with 2xSSC buffer at 55°C. Approximately 30 colonies gave positive hybridization signals with the degenerate probe. Several cosmids were selected and divided into two classes based on restriction digestion patterns. A representative cosmid was selected from each class for further analysis. The
15 representative cosmids were designated pKOS023-26 and pKOS023-27. These cosmids were determined by DNA sequencing to comprise the narbonolide PKS genes, the desosamine biosynthesis and transferase genes, the beta-glucosidase gene, and the *picK* hydroxylase gene.

These cosmids were deposited with the American Type Culture Collection in accordance with the terms of the Budapest Treaty. Cosmid pKOS023-26 was assigned
20 accession number ATCC 203141, and cosmid pKOS023-27 was assigned accession number ATCC 203142.

To demonstrate that the narbonolide PKS genes had been cloned and to illustrate how the invention provides methods and reagents for constructing deletion variants of narbonolide PKS genes, a narbonolide PKS gene was deleted from the chromosome of *Streptomyces*
25 *venezuelae*. This deletion is shown schematically in Figure 4, parts B and C. A ~2.4 kb *EcoRI* - KpnI fragment and a ~2.1 kb KpnI - *XhoI* fragment, which together comprise both ends of the *picA1* gene (but lack a large portion of the coding sequence), were isolated from cosmid pKOS023-27 and ligated together into the commercially available vector pLitmus 28 (digested with restriction enzymes *EcoRI* and *XhoI*) to give plasmid pKOS039-07. The
30 ~4.5 kb *HindIII*-*SpeI* fragment from plasmid pKOS039-07 was ligated with the 2.5 kb *HindIII*-*NheI* fragment of integrating vector pSET152, available from the NRRL, which contains an *E. coli* origin of replication and an apramycin resistance-conferring gene to create

plasmid pKOS039-16. This vector was used to transform *S. venezuelae*, and apramycin-resistant transformants were selected.

Then, to select for double-crossover mutants, the selected transformants were grown in TSB liquid medium without antibiotics for three transfers and then plated onto non-selective media to provide single colony isolates. The isolated colonies were tested for sensitivity to apramycin, and the apramycin-sensitive colonies were then tested to determine if they produced picromycin. The tests performed included a bioassay and LC/MS analysis of the fermentation media. Colonies determined not to produce picromycin (or methymycin or neomethymycin) were then analyzed using PCR to detect an amplification product diagnostic of the deletion. A colony designated K39-03 was identified, providing confirmation that the narbonolide PKS genes had been cloned. Transformation of strain K39-03 with plasmid pKOS039-27 comprising an intact *picA* gene under the control of the *ermE** promoter from plasmid pWHM3 (see Vara *et al.*, *J. Bact.* (1989) 171: 5872-5881, incorporated herein by reference) was able to restore picromycin production.

To determine that the cosmids also contained the *picK* hydroxylase gene, each cosmid was probed by Southern hybridization using a labeled DNA fragment amplified by PCR from the *Saccharopolyspora erythraea* C12-hydroxylase gene, *eryK*. The cosmids were digested with *Bam*HI endonuclease and electrophoresed on a 1% agarose gel, and the resulting fragments were transferred to a nylon membrane. The membrane was incubated with the *eryK* probe overnight at 42°C, washed twice at 25°C in 2XSSC buffer with 0.1% SDS for 15 minutes, followed by two 15 minute washes with 2XSSC buffer at 50°C. Cosmid pKOS023-26 produced an ~3 kb fragment that hybridized with the probe under these conditions. This fragment was subcloned into the PCRscript™ (Stratagene) cloning vector to yield plasmid pKOS023-28 and sequenced. The ~1.2 kb gene designated *picK* above was thus identified. The *picK* gene product is homologous to *eryK* and other known macrolide cytochrome P450 hydroxylases.

By such methodology, the complete set of picromycin biosynthetic genes were isolated and identified. DNA sequencing of the cloned DNA provided further confirmation that the correct genes had been cloned. In addition, and as described in the following example, the identity of the genes was confirmed by expression of narbomycin in heterologous host cells.

Example 3Heterologous Expression of the Narbonolide PKS and the Picromycin Biosynthetic Gene Cluster

To provide a preferred host cell and vector for purposes of the invention, the
5 narbonolide PKS was transferred to the non-macrolide producing host *Streptomyces lividans*
K4-114 (see Ziermann and Betlach, 1999, *Biotechniques* 26, 106-110, and U.S. patent
application Serial No. 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein
by reference). This was accomplished by replacing the three DEBS ORFs on a modified
version of pCK7 (see Kao *et al.*, 1994, *Science* 265, 509-512, and U.S. Patent No. 5,672,491,
10 each of which is incorporated herein by reference) with all four narbonolide PKS ORFs to
generate plasmid pKOS039-86 (see Figure 5). The pCK7 derivative employed, designated
pCK7'Kan', differs from pCK7 only in that it contains a kanamycin resistance conferring
gene inserted at its *HindIII* restriction enzyme recognition site. Because the plasmid contains
two selectable markers, one can select for both markers and so minimize contamination with
15 cells containing rearranged, undesired vectors.

Protoplasts were transformed using standard procedures and transformants selected
using overlays containing antibiotics. The strains were grown in liquid R5 medium for
growth/seed and production cultures at 30°C. A 2 L shake flask culture of *S. lividans* K4-
114/pKOS039-86 was grown for 7 days at 30°C. The mycelia was filtered, and the aqueous
20 layer was extracted with 2 x 2 L ethyl acetate. The organic layers were combined, dried over
MgSO₄, filtered, and evaporated to dryness. Polyketides were separated from the crude
extract by silica gel chromatography (1:4 to 1:2 ethyl acetate:hexane gradient) to give an ~10
mg mixture of narbonolide and 10-deoxymethynolide, as indicated by LC/MS and ¹H NMR.
Purification of these two compounds was achieved by HPLC on a C-18 reverse phase column
25 (20-80% acetonitrile in water over 45 minutes). This procedure yielded ~5 mg each of
narbonolide and 10-deoxymethynolide. Polyketides produced in the host cells were analyzed
by bioassay against *Bacillus subtilis* and by LC/MS analysis. Analysis of extracts by LC/MS
followed by ¹H-NMR spectroscopy of the purified compounds established their identity as
narbonolide (Figure 5, compound 4; see Kaiho *et al.*, 1982, *J. Org. Chem.* 47: 1612-1614,
30 incorporated herein by reference) and 10-deoxymethynolide (Figure 5, compound 5; see
Lambalot *et al.*, 1992, *J. Antibiotics* 45, 1981-1982, incorporated herein by reference), the
respective 14 and 12-membered polyketide aglycones of YC17, narbomycin, picromycin, and
methymycin.

The production of narbonolide in *Streptomyces lividans* represents the expression of an entire modular polyketide pathway in a heterologous host. The combined yields of compounds 4 and 5 are similar to those obtained with expression of DEBS from pCK7 (see Kao *et al.*, 1994, *Science* 265: 509-512, incorporated herein by reference). Furthermore, based on the relative ratios (~1:1) of compounds 4 and 5 produced, it is apparent that the narbonolide PKS itself possesses an inherent ability to produce both 12 and 14-membered macrolactones without the requirement of additional activities unique to *S. venezuelae*. Although the existence of a complementary enzyme present in *S. lividans* that provides this function is possible, it would be unusual to find such a specific enzyme in an organism that does not produce any known macrolide.

To provide a heterologous host cell of the invention that produces the narbonolide PKS and the *picB* gene, the *picB* gene was integrated into the chromosome of *Streptomyces lividans* harboring plasmid pKOS039-86 to yield *S. lividans* K39-18/pKOS039-86. To provide the integrating vector utilized, the *picB* gene was cloned into the *Streptomyces* genome integrating vector pSET152 (see Bierman *et al.*, 1992, *Gene* 116, 43, incorporated herein by reference) under control of the same promoter (*PactI*) as the PKS on plasmid pKOS039-86.

A comparison of strains K39-18/pKOS039-86 and K4-114/pKOS039-86 grown under identical conditions indicated that the strain containing TEII produced 4-7 times more total polyketide. Each strain was grown in 30 mL of R5 (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory Manual*; John Innes Foundation: Norwich, UK, 1985, incorporated herein by reference) liquid (with 20 µg/mL thiostrepton) at 30°C for 9 days. The fermentation broth was analyzed directly by reverse phase HPLC. Absorbance at 235 nm was used to monitor compounds and measure relative abundance. This increased production indicates that the enzyme is functional in this strain. As noted above, because the production levels of compound 4 and 5 from K39-18/pKOS03986 increased by the same relative amounts, TEII does not appear to influence the ratio of 12 and 14-membered lactone ring formation.

To express the glycosylated counterparts of narbonolide (narbomycin) and 10-deoxymethynolide (YC17) in heterologous host cells, the desosamine biosynthetic genes and desosaminyl transferase gene were transformed into the host cells harboring plasmid pKOS039-86 (and, optionally, the *picB* gene, which can be integrated into the chromosome as described above).

Plasmid pKOS039-104, see Figure 6, comprises the desosamine biosynthetic genes, the beta-glucosidase gene, and the desosaminyl transferase gene. This plasmid was constructed by first inserting a polylinker oligonucleotide, containing a restriction enzyme recognition site for *PacI*, a Shine-Dalgarno sequence, and restriction enzyme recognition sites for *NdeI*, *BglII*, and *HindIII*, into a pUC19 derivative, called pKOS24-47, to yield plasmid pKOS039-98.

An ~0.3 kb PCR fragment comprising the coding sequence for the N-terminus of the *desI* gene product and an ~0.12 kb PCR fragment comprising the coding sequence for the C-terminus of the *desR* gene product were amplified from cosmid pKOS23-26 (ATCC 203141) and inserted together into pLitmus28 treated with restriction enzymes *NsiI* and *EcoRI* to produce plasmid pKOS039-101. The ~6 kb *SphI-PstI* restriction fragment of pKOS23-26 containing the *desI*, *desII*, *desIII*, *desIV*, and *desV* genes was inserted into plasmid pUC19 (Stratagene) to yield plasmid pKOS039-102. The ~6 kb *SphI-EcoRI* restriction fragment from plasmid pKOS039-102 was inserted into pKOS039-101 to produce plasmid pKOS039-103. The ~6 kb *BglII-PstI* fragment from pKOS23-26 that contains the *desR*, *desVI*, *desVII*, and *desVIII* genes was inserted into pKOS39-98 to yield pKOS39-100. The ~6 kb *PacI-PstI* restriction fragment of pKOS39-100 and the ~6.4 kb *NsiI-EcoRI* fragment of pKOS39-103 were cloned into pKOS39-44 to yield pKOS39-104.

When introduced into *Streptomyces lividans* host cells comprising the recombinant narbonolide PKS of the invention, plasmid pKOS39-104 drives expression of the desosamine biosynthetic genes, the beta-glucosidase gene, and the desosaminyl transferase gene. The glycosylated antibiotic narbomycin was produced in these host cells, and it is believed that YC17 was produced as well. When these host cells are transformed with vectors that drive expression of the *picK* gene, the antibiotics methymycin, neomethymycin, and picromycin are produced.

In similar fashion, when plasmid pKOS039-18, which encodes a hybrid PKS of the invention that produces 3-deoxy-3-oxo-6-deoxyerythronolide B was expressed in *Streptomyces lividans* host cells transformed with plasmid pKOS39-104, the 5-desosaminylated analog was produced. Likewise, when plasmid pCK7, which encodes DEBS, which produces 6-deoxyerythronolide B, was expressed in *Streptomyces lividans* host cells transformed with plasmid pKOS39-104, the 5-desosaminylated analog was produced. These compounds have antibiotic activity and are useful as intermediates in the synthesis of other antibiotics.

Example 4

Expression Vector for Desosaminyl Transferase

While the invention provides expression vectors comprising all of the genes required
5 for desosamine biosynthesis and transfer to a polyketide, the invention also provides
expression vectors that encode any subset of those genes or any single gene. As one
illustrative example, the invention provides an expression vector for desosaminyl transferase.
This vector is useful to desosaminylate polyketides in host cells that produce NDP-
desosamine but lack a desosaminyl transferase gene or express a desosaminyl transferase that
10 does not function as efficiently on the polyketide of interest as does the desosaminyl
transferase of *Streptomyces venezuelae*. This expression vector was constructed by first
amplifying the desosaminyl transferase coding sequence from pKOS023-27 using the
primers:
N3917: 5'-CCCTGCAGCGGCAAGGAAGGACACGACGCCA-3' (SEQ ID NO:25); and
15 N3918: 5'-AGGTCTAGAGCTCAGTGCCGGGCGTCGGCCGG-3' (SEQ ID NO:26),
to give a 1.5 kb product. This product was then treated with restriction enzymes *Pst*I and
*Xba*I and ligated with *Hind*III and *Xba*I digested plasmid pKOS039-06 together with the 7.6
kb *Pst*I-*Hind*III restriction fragment of plasmid pWHM1104 to provide plasmid pKOS039-
14. Plasmid pWHM1104, described in Tang *et al.*, 1996, *Molec. Microbiol.* 22(5): 801-813,
20 incorporated herein by reference, encodes the *ermE** promoter. Plasmid pKOS039-14 is
constructed so that the desosaminyl transferase gene is placed under the control of the *ermE**
promoter and is suitable for expression of the desosaminyl transferase in *Streptomyces*,
Saccharopolyspora erythraea, and other host cells in which the *ermE** promoter functions.

25

Example 5

Heterologous Expression of the *picK* Gene Product in *E. coli*

The *picK* gene was PCR amplified from plasmid pKOS023-28 using the
oligonucleotide primers:
N024-36B (forward):
30 5'-TTGCATGCATATGCGCCGTACCCAGCAGGGAACGACC (SEQ ID NO:27); and
N024-37B (reverse):
5'-TTGAATTCTCAACTAGTACGGCGGCCCGCCTCCCGTCC (SEQ ID NO:28). These
primers alter the *Streptomyces* GTG start codon to ATG and introduce a *Spe*I site at the C-

terminal end of the gene, resulting in the substitution of a serine for the terminal glycine amino acid residue. The blunt-ended PCR product was subcloned into the commercially available vector pCRscript at the SrfI site to yield plasmid pKOS023-60. An ~1.3 kb *NdeI*-*XhoI* fragment was then inserted into the *NdeI/XhoI* sites of the T7 expression vector pET22b (Novagen, Madison, WI) to generate pKOS023-61. Plasmid pKOS023-61 was digested with restriction enzymes *SpeI* and *EcoRI*, and a short linker fragment encoding 6 histidine residues and a stop codon (composed of oligonucleotides 30-85a: 5'-CTAGTATGCATCATCATCATCATTA-3' (SEQ ID NO:29); and 30-85b: 5'-AATTTTAATGATGATGATGATGATGCATA-3' (SEQ ID NO:30) was inserted to obtain plasmid pKOS023-68. Both plasmid pKOS023-61 and pKOS023-68 produced active PicK enzyme in recombinant *E. coli* host cells.

Plasmid pKOS023-61 was transformed into *E. coli* BL21-DE3. Successful transformants were grown in LB-containing carbenicillin (100 µg/ml) at 37°C to an OD600 of 0.6. Isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and the cells were grown for an additional 3 hours before harvesting. The cells were collected by centrifugation and frozen at -80°C. A control culture of BL21-DE3 containing the vector plasmid pET21c (Invitrogen) was prepared in parallel.

The frozen BL21-DE3/pKOS023-61 cells were thawed, suspended in 2 µL of cold cell disruption buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris/HCl, pH 8.0) and sonicated to facilitate lysis. Cellular debris and supernatant were separated by centrifugation and subjected to SDS-PAGE on 10-15% gradient gels, with Coomassie Blue staining, using a Pharmacia Phast Gel Electrophoresis system. The soluble crude extract from BL21-DE3/pKOS023-61 contained a Coomassie stained band of Mr~46 kDa, which was absent in the control strain BL21-DE3/pET21c.

The hydroxylase activity of the *picK* protein was assayed as follows. The crude supernatant (20 µL) was added to a reaction mixture (100 µL total volume) containing 50 mM Tris/HCl (pH 7.5), 20 µM spinach ferredoxin, 0.025 Unit of spinach ferredoxin:NADP+ oxidoreductase, 0.8 Unit of glucose-6-phosphate dehydrogenase, 1.4 mM NADP+, 7.6 mM glucose-6-phosphate, and 20 nmol of narbomycin. The narbomycin was purified from a culture of *Streptomyces narbonensis*, and upon LC/MS analysis gave a single peak of [M+H]⁺=510. The reaction was allowed to proceed for 105 minutes at 30°C. Half of the reaction mixture was loaded onto an HPLC, and the effluent was analyzed by evaporative light scattering (ELSD) and mass spectrometry. The control extract (BL21-DE3/pET21c) was

processed identically. The BL21-DE3/pKOS023-61 reaction contained a compound not present in the control having the same retention time, molecular weight and mass fragmentation pattern as picromycin ($[M+H]^+=526$). The conversion of narbomycin to picromycin under these conditions was estimated to be greater than 90% by ELSD peak area.

5 The poly-histidine-linked PicK hydroxylase was prepared from pKOS023-68 transformed into *E. coli* BL21 (DE3) and cultured as described above. The cells were harvested and the PicK protein purified as follows. All purification steps were performed at 4°C. *E. coli* cell pellets were suspended in 32 μ L of cold binding buffer (20 mM Tris/HCl, pH 8.0, 5 mM imidazole, 500 mM NaCl) per mL of culture and lysed by sonication. For
10 analysis of *E. coli* cell-free extracts, the cellular debris was removed by low-speed centrifugation, and the supernatant was used directly in assays. For purification of PicK/6-His, the supernatant was loaded (0.5 mL/min.) onto a 5 mL HiTrap Chelating column (Pharmacia, Piscataway, New Jersey), equilibrated with binding buffer. The column was washed with 25 μ L of binding buffer and the protein was eluted with a 35 μ L linear gradient
15 (5-500 mM imidazole in binding buffer). Column effluent was monitored at 280 nm and 416 nm. Fractions corresponding to the 416 nm absorbance peak were pooled and dialyzed against storage buffer (45 mM Tris/HCl, pH 7.5, 0.1 mM EDTA, 0.2 mM DTT, 10% glycerol). The purified 46 kDa protein was analyzed by SDS-PAGE using Coomassie blue staining, and enzyme concentration and yield were determined.

20 Narbomycin was purified as described above from a culture of *Streptomyces narbonensis* ATCC19790. Reactions for kinetic assays (100 μ L) consisted of 50 mM Tris/HCl (pH 7.5), 100 μ M spinach ferredoxin, 0.025 Unit of spinach ferredoxin:NADP⁺ oxidoreductase, 0.8 U glucose-6-phosphate dehydrogenase, 1.4 mM NADP⁺, 7.6 mM glucose-6-phosphate, 20-500 μ M narbomycin substrate, and 50-500 nM of PicK enzyme. The
25 reaction proceeded at 30°C, and samples were withdrawn for analysis at 5, 10, 15, and 90 minutes. Reactions were stopped by heating to 100°C for 1 minute, and denatured protein was removed by centrifugation. Depletion of narbomycin and formation of picromycin were determined by high performance liquid chromatography (HPLC, Beckman C-18 0.46x15 cm column) coupled to atmospheric pressure chemical ionization (APCI) mass spectroscopic
30 detection (Perkin Elmer/Sciex API 100) and evaporative light scattering detection (Alltech 500 ELSD).

Example 6

Expression of the *picK* Gene Encoding the Hydroxylase in *Streptomyces narbonensis*

To produce picromycin in *Streptomyces narbonensis*, a host that produces narbomycin but not picromycin, the methods and vectors of the invention were used to
5 express the *picK* gene in this host.

The *picK* gene was amplified from cosmid pKOS023-26 using the primers:

N3903: 5'-TCCTCTAGACGTTTCCGT-3' (SEQ ID NO:31); and

N3904: 5'-TGAAGCTTGAATTCAACCGGT-3' (SEQ ID NO:32)

to obtain an ~1.3 kb product. The product was treated with restriction enzymes *XbaI* and
10 *HindIII* and ligated with the 7.6 kb *XbaI-HindIII* restriction fragment of plasmid pWHM1104 to provide plasmid pKOS039-01, placing the *picK* gene under the control of the *ermE** promoter. The resulting plasmid was transformed into purified stocks of *S. narbonensis* by protoplast fusion and electroporation. The transformants were grown in suitable media and shown to convert narbomycin to picromycin at a yield of over 95%.

15

Example 7

Construction of a Hybrid DEBS/Narbonolide PKS

This example describes the construction of illustrative hybrid PKS expression vectors of the invention. The hybrid PKS contains portions of the narbonolide PKS and portions of
20 rapamycin and/or DEBS PKS. In the first constructs, pKOS039-18 and pKOS039-19, the hybrid PKS comprises the narbonolide PKS extender module 6 ACP and thioesterase domains and the DEBS loading module and extender modules 1-5 as well as the KS and AT domains of DEBS extender module 6 (but not the KR domain of extender module 6). In pKOS039-19, the hybrid PKS is identical except that the KS1 domain is inactivated, i.e., the
25 ketosynthase in extender module 1 is disabled. The inactive DEBS KS1 domain and its construction are described in detail in PCT publication Nos. WO 97/02358 and WO 99/03986, each of which is incorporated herein by reference. To construct pKOS039-18, the 2.33 kb *BamHI-EcoRI* fragment of pKOS023-27, which contains the desired sequence, was amplified by PCR and subcloned into plasmid pUC19. The primers used in the PCR
30 were:

N3905: 5'-TTTATGCATCCCGCGGGTCCCGGCGAG-3' (SEQ ID NO:33); and

N3906: 5'-TCAGAATTCTGTCGGTCACTTGCCCGC-3' (SEQ ID NO:34).

The 1.6 kb PCR product was digested with *Pst*I and *Eco*RI and cloned into the corresponding sites of plasmid pKOS015-52 (this plasmid contains the relevant portions of the coding sequence for the DEBS extender module 6) and commercially available plasmid pLitmus 28 to provide plasmids pKOS039-12 and pKOS039-13, respectively. The BglII - *Eco*RI
5 fragment of plasmid pKOS039-12 was cloned into plasmid pKOS011-77, which contains the functional DEBS gene cluster and into plasmid pJRJ2, which contains the mutated DEBS gene that produces a DEBS PKS in which the KS domain of extender module I has been rendered inactive. Plasmid pJRJ2 is described in PCT publication Nos. WO 99/03986 and WO 97/02358, incorporated herein by reference.

10 Plasmids pKOS039-18 and pKOS039-19, respectively, were obtained. These two plasmids were transformed into *Streptomyces coelicolor* CH999 by protoplast fusion. The resulting cells were cultured under conditions such that expression of the PKS occurred. Cells transformed with plasmid pKOS039-18 produced the expected product 3-deoxy-3-oxo-6-deoxyerythronolide B. When cells transformed with plasmid pKOS039-19 were provided
15 (2S,3R)-2-methyl-3-hydroxyhexanoate NACS, 13-desethyl-13-propyl-3-deoxy-3-oxo-6-deoxyerythronolide B was produced.

Example 8

6-Hydroxylation of 3,6-dideoxy-3-oxoerythronolide B using the eryF hydroxylase

20 Certain compounds of the invention can be hydroxylated at the C6 position in a host cell that expresses the eryF gene. These compounds can also be hydroxylated *in vitro*, as illustrated by this example.

The 6-hydroxylase encoded by eryF was expressed in *E. coli*, and partially purified. The hydroxylase (100 pmol in 10 μ L) was added to a reaction mixture (100 μ l total volume)
25 containing 50 mM Tris/HCl (pH 7.5), 20 μ M spinach ferredoxin, 0.025 Unit of spinach ferredoxin:NADP+ oxidoreductase, 0.8 Unit of glucose-6-phosphate dehydrogenase, 1.4 mM NADP+, 7.6 mM glucose-6-phosphate, and 10 nmol 6-deoxyerythronolide B. The reaction was allowed to proceed for 90 minutes at 30°C. Half of the reaction mixture was loaded onto an HPLC, and the effluent was analyzed by mass spectrometry. The production of
30 erythronolide B as evidenced by a new peak eluting earlier in the gradient and showing [M+H]⁺=401. Conversion was estimated at 50% based on relative total ion counts.

Those of skill in the art will recognize the potential for hemiketal formation in the above compound and compounds of similar structure. To reduce the amount of hemiketal

formed, one can use more basic (as opposed to acidic) conditions or employ sterically hindered derivative compounds, such as 5-desosaminylated compounds.

Example 9

5

Measurement of Antibacterial Activity

Antibacterial activity was determined using either disk diffusion assays with *Bacillus cereus* as the test organism or by measurement of minimum inhibitory concentrations (MIC) in liquid culture against sensitive and resistant strains of *Staphylococcus pneumoniae*.

10 The invention having now been described by way of written description and example, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments and that the foregoing description and examples are for purposes of illustration and not limitation of the following claims.

Claims

1. A recombinant DNA compound that comprises a coding sequence for a domain of a narbonolide PKS.
5
2. The recombinant DNA compound of claim 1, wherein said domain is selected from the group consisting of a thioesterase domain, a KS^Q domain, an AT domain, a KS domain, an ACP domain, a KR domain, a DH domain, and an ER domain.
- 10 3. The recombinant DNA compound of claim 2 that comprises the coding sequence for a loading module, thioesterase domain, and all six extender modules of the narbonolide PKS.
4. The recombinant DNA compound of claim 2 that comprises a hybrid PKS.
15
5. The recombinant DNA compound of claim 4 wherein said hybrid PKS comprises at least a portion of a narbonolide PKS gene and at least a portion of a second PKS gene for a macrolide aglycone other than narbonolide.
- 20 6. The recombinant DNA compound of claim 5 wherein said second PKS gene is a DEBS gene.
7. The recombinant DNA compound of claim 6 wherein said hybrid PKS is composed of a loading module and extender modules 1 through 6 of DEBS excluding a KR domain of extender module 6 of DEBS and an ACP of extender module 6 and a thioesterase domain of the narbonolide PKS.
25
8. A recombinant DNA compound that comprises a coding sequence for a desosamine biosynthetic gene or a desosaminyl transferase gene or a beta-glucosidase gene of
30 *Streptomyces venezuelae*.
9. A recombinant DNA compound that comprises a coding sequence for a *pick* hydroxylase gene of *Streptomyces venezuelae*.

10. The DNA compound of any of claims 1-9 that further comprises a promoter operably linked to said coding sequence.

5 11. The recombinant DNA compound of claim 10, wherein said promoter is a promoter derived from a cell other than a *Streptomyces venezuelae* cell.

12. The recombinant DNA compound of claim 11 that is a recombinant DNA expression vector.

10

13. The expression vector of claim 12 that expresses a PKS in *Streptomyces* host cells.

14. A recombinant host cell, which in its untransformed state does not produce 10-deoxymethynolide or narbonolide, that comprises a recombinant DNA expression vector of claim 12 that encodes a narbonolide PKS and said cell produces 10-deoxymethynolide or narbonolide.

15. The recombinant host cell of claim 14 that further comprises a *picB* gene.

20

16. The recombinant host cell of claim 14 that further comprises desosamine biosynthetic genes and a gene for desosaminyl transferase and produces YC17 or narbomycin.

25 17. The recombinant host cell of claim 16 that further comprises a *picK* gene and produces methymycin, neomethymycin, or picromycin.

18. The recombinant host cell of any of claim 17 that is *Streptomyces coelicolor* or *Streptomyces lividans*.

30

19. A recombinant host cell other than a *Streptomyces venezuelae* cell that expresses the *picK* hydroxylase gene of *S. venezuelae*.

20. A recombinant host cell other than a *Streptomyces venezuelae* host cell that expresses a desosamine biosynthetic gene or desosaminyl transferase gene of *S. venezuelae*.

21. A method for increasing the yield of a desosaminylated polyketide in a cell,
5 which method comprises transforming the cell with a recombinant expression vector that encodes a functional beta-glucosidase gene.

22. A hybrid PKS which comprises at least one domain of a narbonolide PKS.

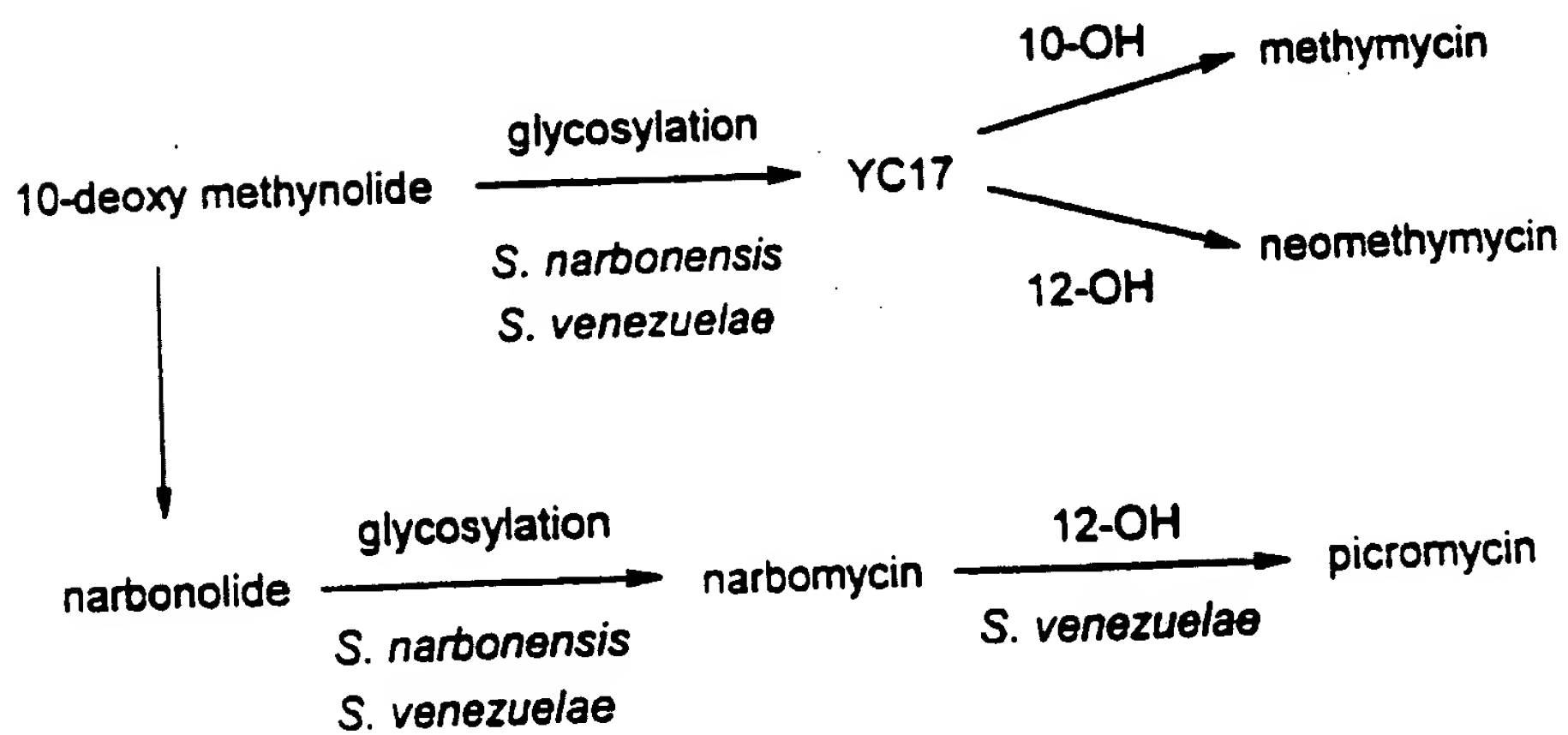
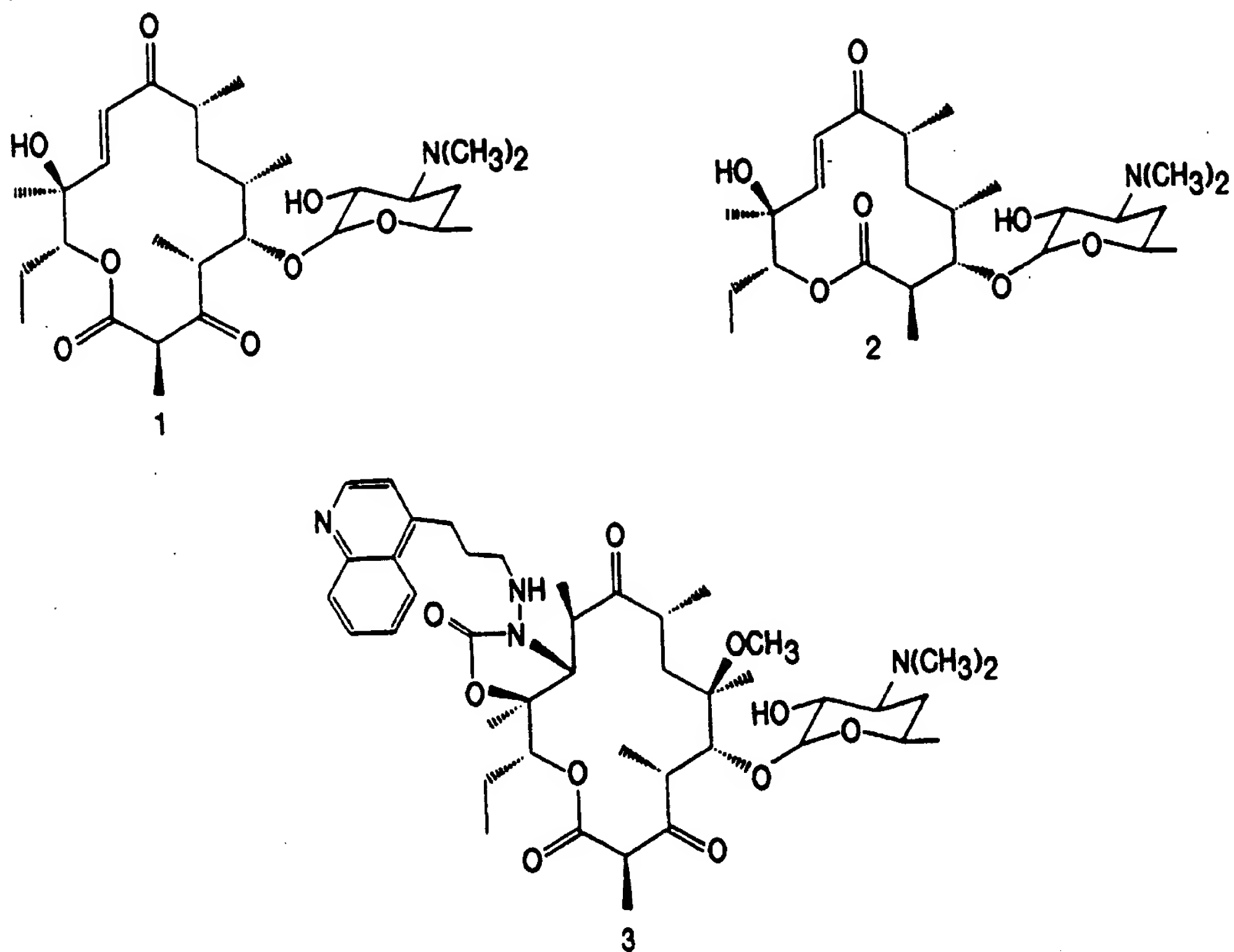
10 23. The hybrid PKS of claim 22 wherein said hybrid PKS comprises at least a portion of a narbonolide PKS gene and at least a portion of a second PKS gene for a macrolide aglycone other than narbonolide.

15 24. The hybrid PKS of claim 23 wherein said second PKS gene is a DEBS gene.

25. The hybrid PKS of claim 24 wherein said hybrid PKS is composed of a loading module and extender modules 1 through 6 of DEBS excluding a KR domain of extender module 6 of DEBS and an ACP of extender module 6 and a thioesterase domain of the narbonolide PKS.

20 26. A method to produce a polyketide which comprises providing starter, extender and/or intermediate ketide units to the hybrid PKS of claim 22.

25 27. A polyketide produced by the method of claim 26.



5

FIG. 1

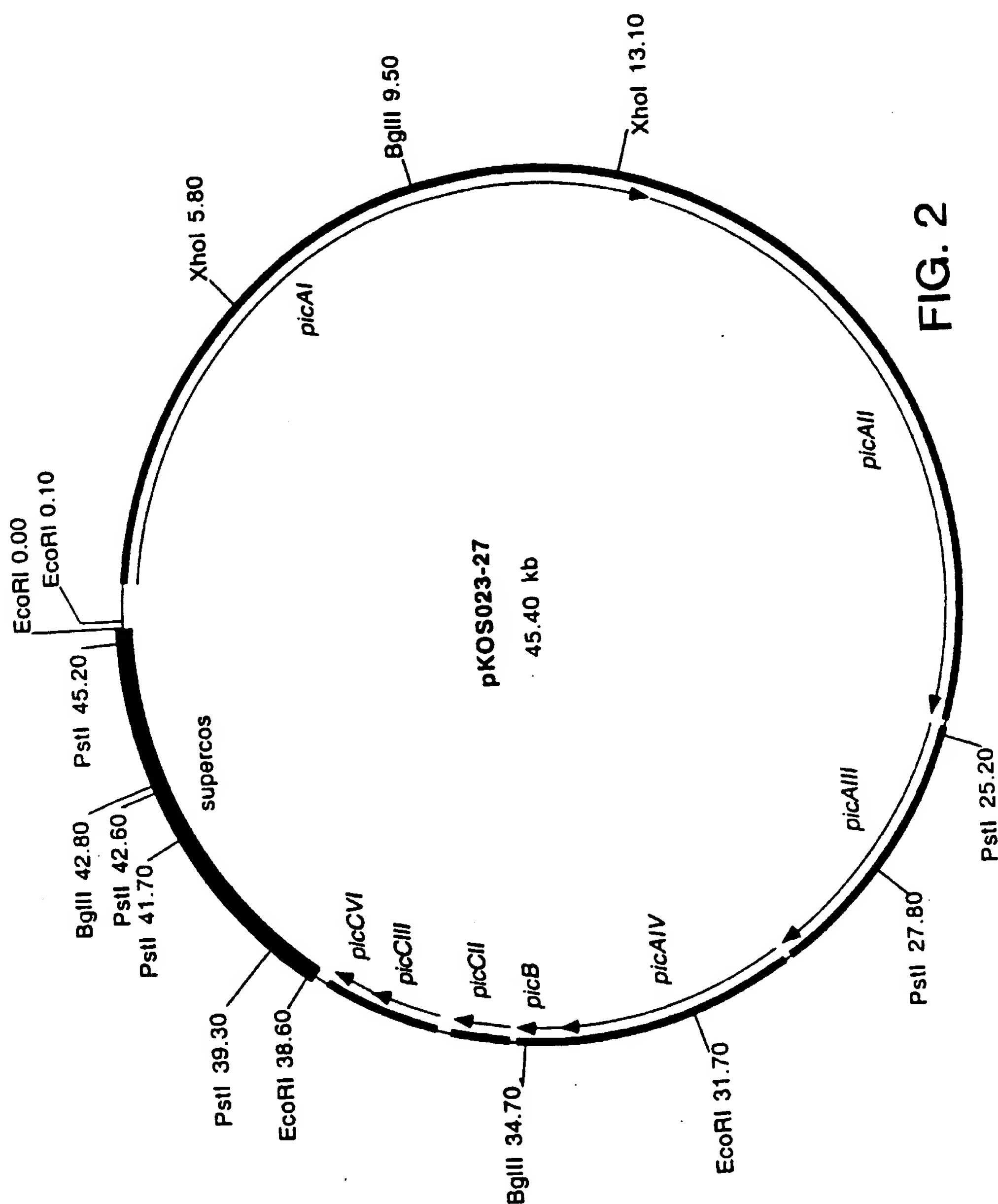


FIG. 2

3/7

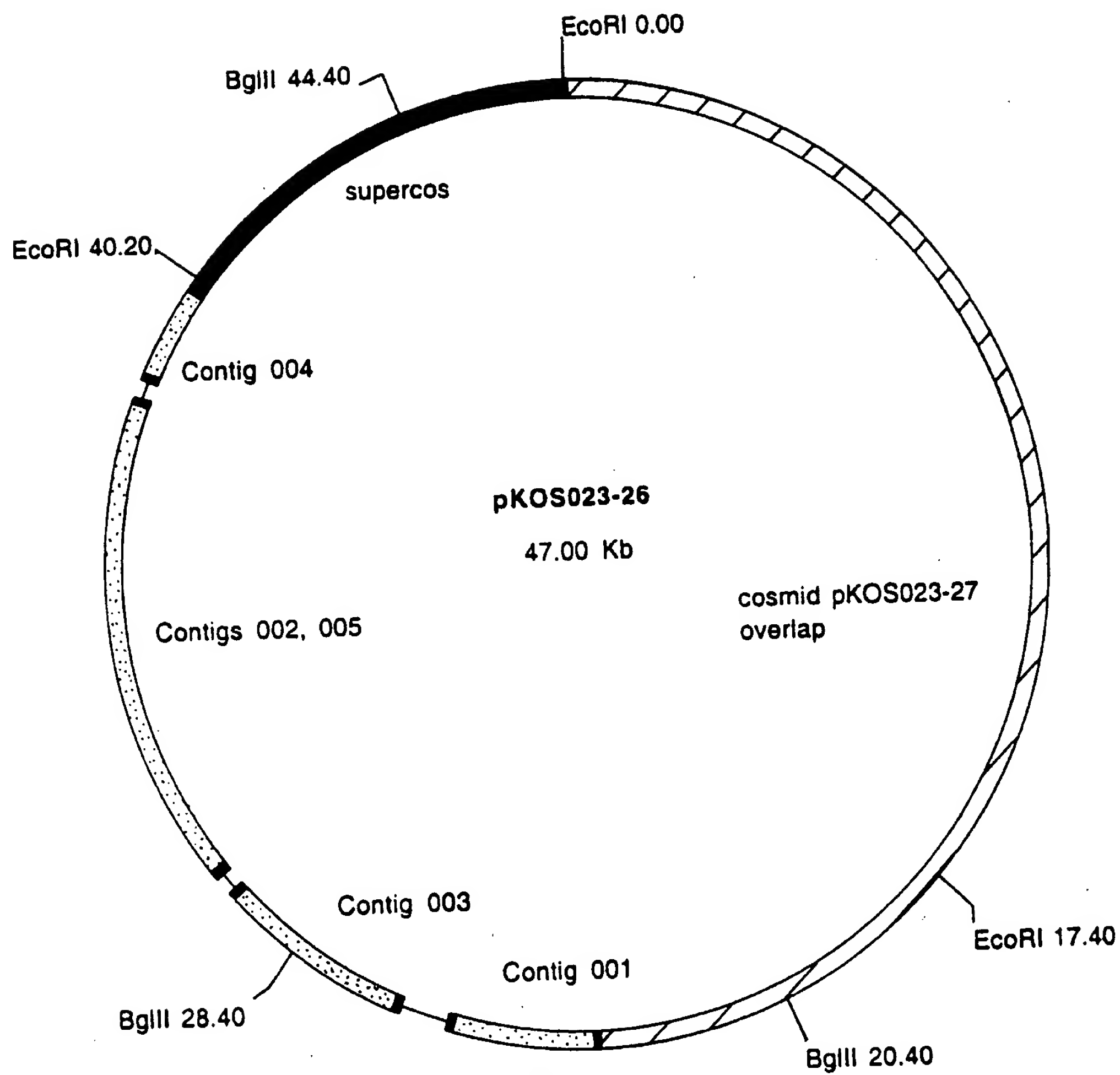


FIG. 3

4/7

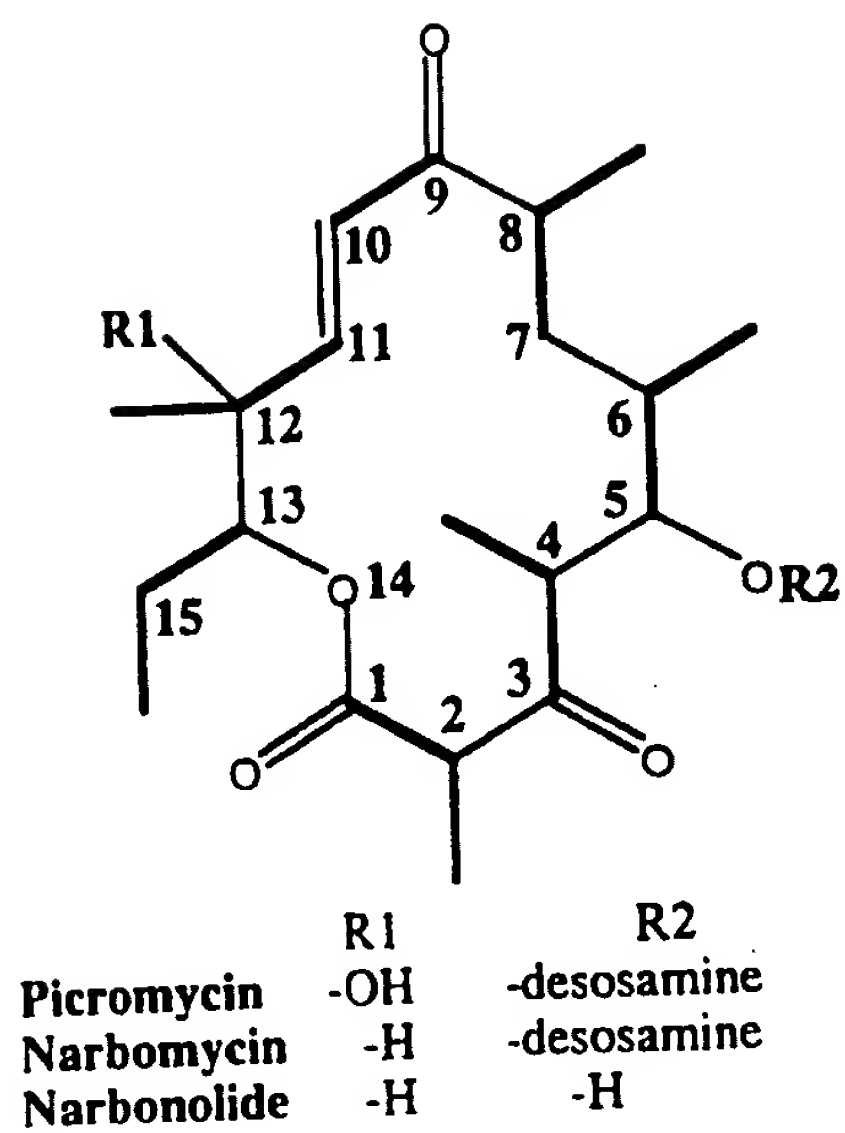


FIG. 4A-1

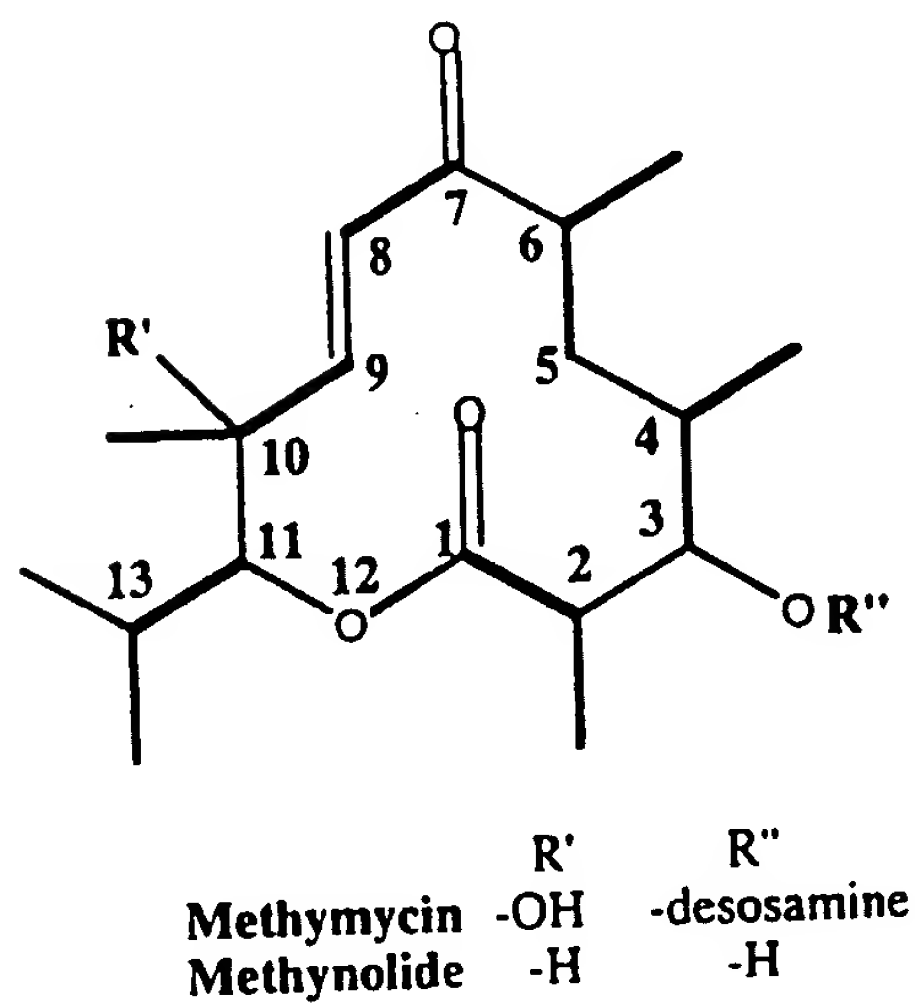
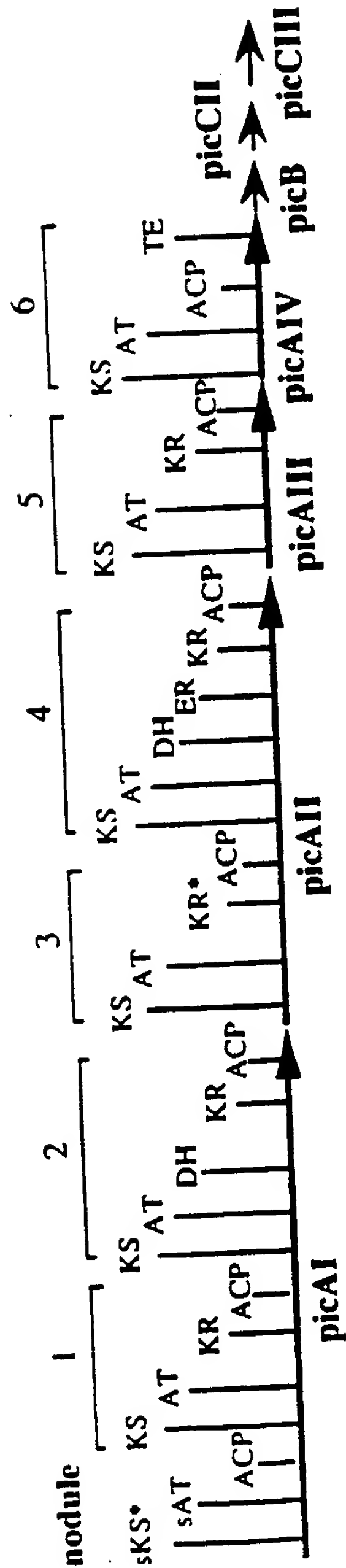


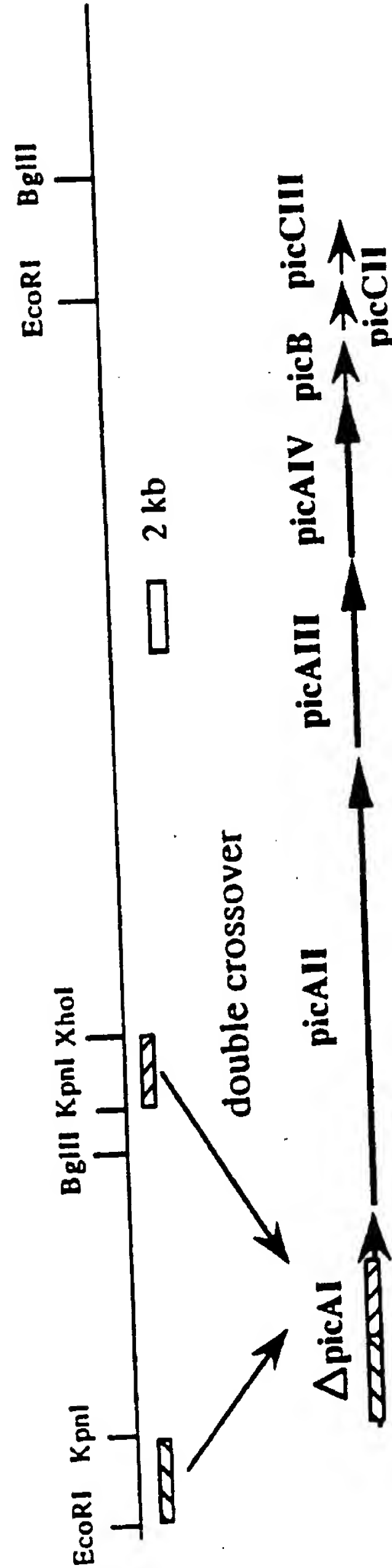
FIG. 4A-2

5/7



S. venezuelae
ATCC15439

FIG. 4B



pKOS023-27

S. venezuelae
K039-03

FIG. 4C

6/7

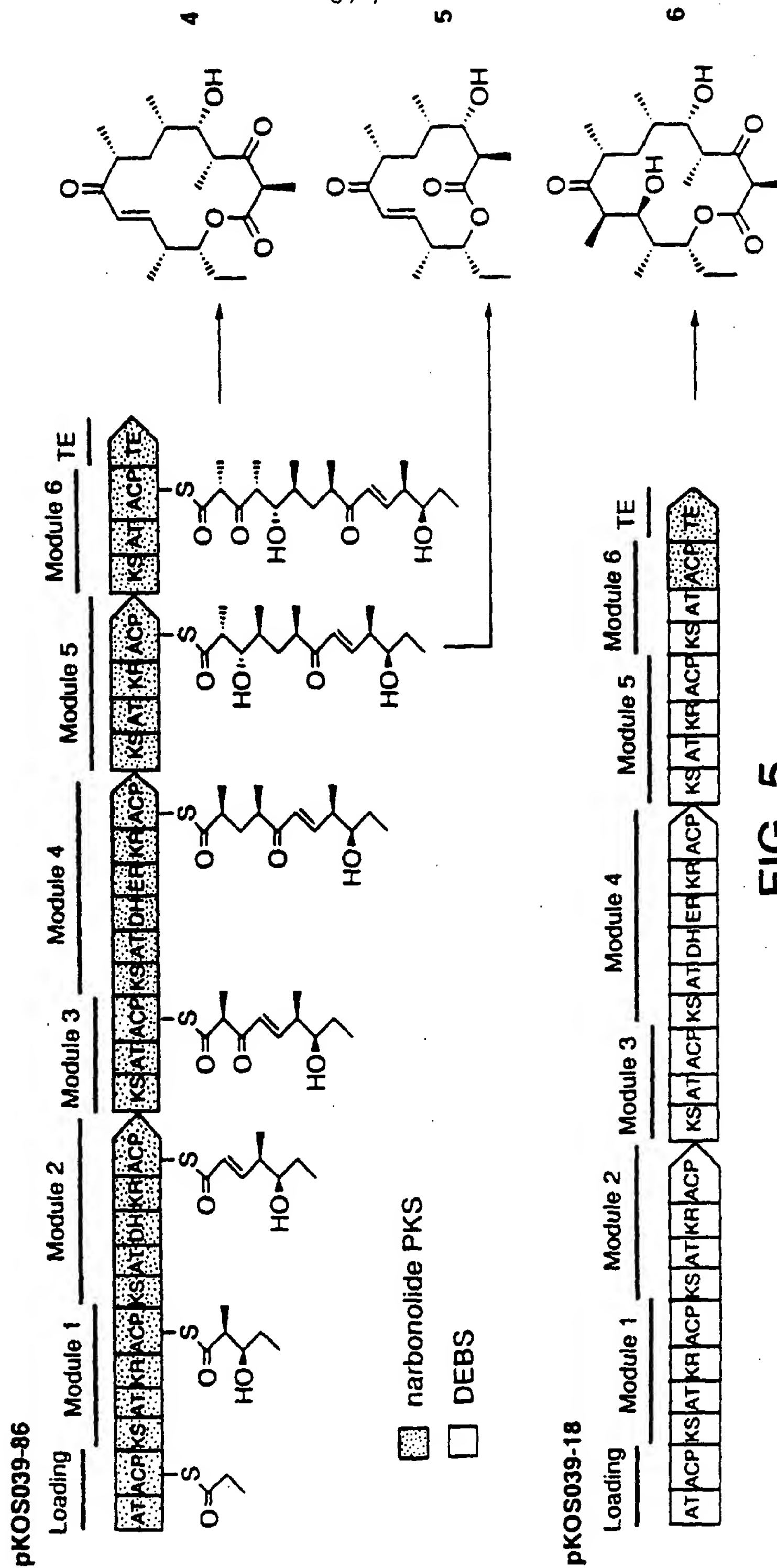


FIG. 5

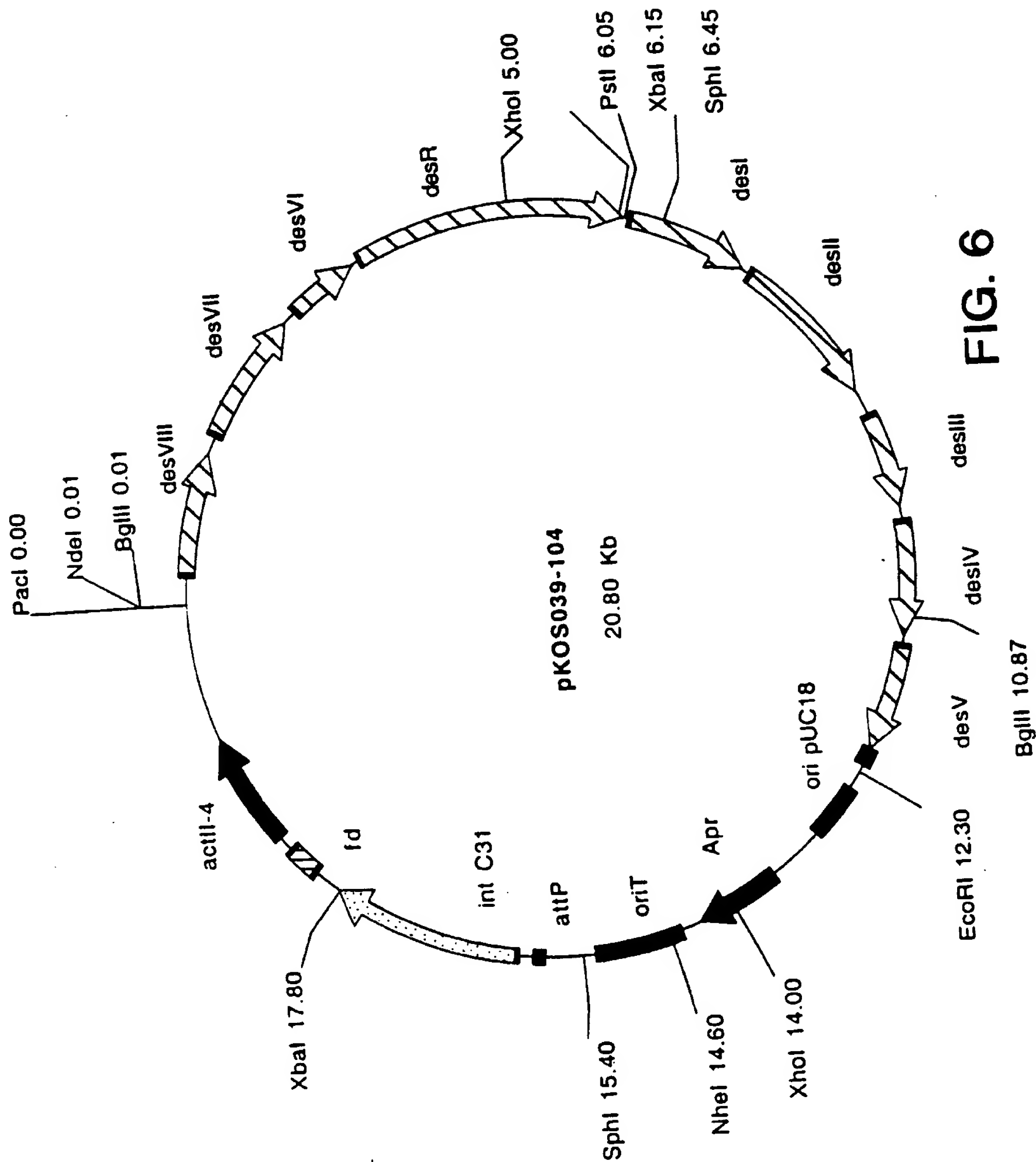


FIG. 6

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/11814

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/52 C12N15/76 C12N9/00 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 40968 A (UNIV LELAND STANFORD JUNIOR ;JOHN INNES CENTRE (GB)) 19 December 1996 (1996-12-19) the whole document ---	1
A	WO 98 01546 A (CORTES JESUS ;LEADLAY PETER F (GB); STAUNTON JAMES (GB); BIOTICA T) 15 January 1998 (1998-01-15) the whole document --- -/--	1

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

3 December 1999

Date of mailing of the international search report

10/12/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.
Fax: (+31-70) 340-3016

Authorized officer

Hillenbrand, G

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/11814

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>BROWN M J B ET AL: "A MUTANT GENERATED BY EXPRESSION OF AN ENGINEERED DEBS1 PROTEIN FROM THE ERYTHROMYCIN-PRODUCING POLYKETIDE SYNTHASE (PKS) IN STREPTOMYCES COELICOLOR PRODUCES THE TRIKETIDE AS A LACTONE, BUT THE MAJOR PRODUCT IS THE NOR-ANALOGUE DERIVED FROM ACETATE AS STARTER ACID" JOURNAL OF THE CHEMICAL SOCIETY, CHEMICAL COMMUNICATIONS, GB, CHEMICAL SOCIETY. LETCHWORTH, no. 15, 1995, page 1517-1518 XP002044729 ISSN: 0022-4936 the whole document</p>	1

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/11814

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9640968 A	19-12-1996	US 5712146 A	27-01-1998
		AU 703920 B	01-04-1999
		AU 6157596 A	30-12-1996
		CA 2224104 A	19-12-1996
		EP 0871760 A	21-10-1998
		NZ 310729 A	29-09-1999
		US 5962290 A	05-10-1999
WO 9801546 A	15-01-1998	AU 3450997 A	02-02-1998
		AU 3451497 A	02-02-1998
		CA 2259420 A	15-01-1998
		CA 2259463 A	15-01-1998
		CN 1229438 A	22-09-1999
		EP 0909327 A	21-04-1999
		EP 0910633 A	28-04-1999
		WO 9801571 A	15-01-1998
		GB 2331518 A	26-05-1999
		NO 990012 A	23-02-1999
		PL 331285 A	05-07-1999
		AU 7666198 A	30-12-1998
		WO 9854308 A	03-12-1998